CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 127, No. 3, Pt. 1 of 2 March 2010

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

NONLINEAR ACOUSTICS [25]

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

The use of interaural time and level difference cues by bilateral cochlear implant users

Justin M. Aronoff,[a] Yang-soo Yoon, and Daniel J. Freed[b]
Communication and Neuroscience Division, House Ear Institute, 2100 W. 3rd St., Los Angeles, California 90057
jaronoff@hei.org, yyoon@hei.org, dfreed@earlenscorp.com

Andrew J. Vermiglio
Department of Communication Disorders and Sciences, California State University, 18111 Nordhoff Street, Northridge, California 91330
andrew.vermiglio@csun.edu

Ivan Pal
Compreval, Inc., 3 Spinnaker Court, Grayslake, Illinois 60030
ivan.compreval@gmail.com

Sigfrid D. Soli
Communication and Neuroscience Division, House Ear Institute, 2100 W. 3rd St., Los Angeles, California 90057
ssoli@hei.org

Abstract: While considerable evidence suggests that bilateral cochlear implant (CI) users' sound localization abilities rely primarily on interaural level difference (ILD) cues, and only secondarily, if at all, on interaural time difference (ITD) cues, this evidence has largely been indirect. This study used head-related transfer functions (HRTFs) to independently manipulate ITD and ILD cues and directly measure their contribution to bilateral CI users' localization abilities. The results revealed a strong reliance on ILD cues, but some CI users also made use of ITD cues. The results also suggest a complex interaction between ITD and ILD cues.

PACS numbers: 43.66.Pn, 43.66.Qp, 43.66.Ts [JL]
Date Received: November 18, 2009 **Date Accepted:** December 24, 2009

1. Introduction

Bilateral cochlear implant (CI) users demonstrate a significant bilateral benefit for localization (e.g., van Hoesel, 2004). Although considerable evidence suggests that this localization benefit is dominated by interaural level difference (ILD) cues, as opposed to interaural time difference (ITD) cues (van Hoesel and Tyler, 2003; Schoen *et al.*, 2005; Grantham *et al.*, 2007; Poon *et al.*, 2009), this evidence has largely been indirect. Some studies have only minimized either the ITD or the ILD cues rather than completely eliminating them (Schoen *et al.*, 2005; Grantham *et al.*, 2007). Other studies have measured ITD sensitivity in μs and ILD sensitivity in dB (van Hoesel, 2004; Schoen *et al.*, 2005; Poon *et al.*, 2009) rather than using a secondary measure, such as spatial resolution, that would allow for a direct comparison between the two.

Recently, direct measurements of the relative contribution of ITD and ILD cues were performed by Seeber and Fastl (2008). They used a single participant's own head-related transfer function (HRTF), allowing them to create stimuli that contained the ITD cues for one location and the ILD cues for a separate location. Consequently, they were able to measure the

[a]Author to whom correspondence should be addressed.
[b]Presently at: EarLens Corp., 200 Chesapeake Dr., Redwood City, CA 94063.

Table 1. Participant information.

Participant (gender, age)	Etiology (patient report)	Processor		Length of use (years)	
		Left	Right	Left	Right
FF1 (F, 62)[a]	Ototoxic	Cochlear Freedom	Cochlear Freedom	2	2
FF2 (F, 42)	Ototoxic	Cochlear Freedom	Cochlear Freedom	0.5	0.5
TT1 (M, 42)[a]	Unknown	Med-El Tempo+	Med-El Tempo+	6	6
TT2 (M, 63)[a]	Noise exposure	Med-El Tempo+	Med-El Tempo+	7	7
TT3 (F, 47)	Unknown	Med-El Tempo+	Med-El Tempo+	6	6
OO1 (M, 58)	Mumps/Measles	Med-El Opus 2	Med-El Opus 2	1	1[b]
OO2 (M, 68)	Ménière's	Med-El Opus 2	Med-El Opus 2	2	0.5
OT1 (F, 37)[a]	Familial	Med-El Opus 2	Med-El Tempo+	3	5
TO1 (M, 51)	Ménière's	Med-El Tempo+	Med-El Opus 2	7	1
TO2 (F, 71)	Unknown	Med-El Tempo+	Med-El Opus 2	7	3

[a]These participants only participated in the SF and DC ITD+ILD conditions.
[b]This participant originally was implanted on the right side 3 years before testing, but electrodes started turning on and off unpredictably within 3 months, and the device was replaced approximately 1 year before testing.

availability of each cue using a single measure of performance, the perceived location. Their results demonstrated that ILD cues had a primary role in localization, with no evidence of a secondary role for ITD cues.

One limitation of the Seeber and Fastl (2008) study is that they used stimuli with the ITD cues for one location and the ILD cues for another rather than presenting each cue in isolation. Given that many studies have indicated a strong dominance for ILD cues (van Hoesel and Tyler, 2003; Schoen *et al.*, 2005; Grantham *et al.*, 2007; Poon *et al.*, 2009), any effect the ITD cues might have had on the perceived location in the Seeber and Fastl (2008) study may have been masked by the more dominant ILD cues. Another limitation of the Seeber and Fastl (2008) study is that only a single subject was tested, making it difficult to generalize their findings, especially since other studies have shown considerable variability in ITD sensitivity across CI users (e.g., van Hoesel, 2004; Poon *et al.*, 2009).

The purpose of the present study was to measure bilateral CI users' ability to use ITD cues and determine how this compared and related to their ability to use ILD cues for sound localization. Because HRTFs were used to independently manipulate ITD and ILD cues, the first goal was to verify that the HRTFs appropriately simulated the soundfield (SF). Given that other researchers have demonstrated considerable variability across bilateral CI users in terms of ITD sensitivity (e.g., Poon *et al.*, 2009), the second goal was to determine if the primary role of ILD cues in localization, reported in Seeber and Fastl (2008) based on one subject, extends to a larger population of bilateral CI users. The third goal was to determine if ITD cues played a secondary role in localization.

2. Methods

Six bilateral CI users participated in a localization task both in the SF and with a process that simulates the microphone input by sending HRTF-processed signals through a cable directly connected to the CI processor's auxiliary input port (Chan *et al.*, 2008), referred to as direct connect (DC) testing. For DC testing, there were three conditions: (1) both ITD and ILD cues present (ITD+ILD), (2) only ITD cues present (ITD-only), and (3) only ILD cues present (ILD-only). Four additional bilateral CI users completed only the ITD+ILD condition as well as SF testing (see Table 1 for listener details). Participants were assigned identifiers using the following con-

vention: [Left processor] [Right processor] [sequential number], where the processors were coded as F = Cochlear's Freedom, O = Med-El's Opus 2, and T = Med-El's Tempo+ (e.g., the subject identifier FF2 represents the second bilateral Freedom processor user tested).

Generic HRTFs were created from recordings from each processor's microphone with the processor positioned on a Knowles Electronics Manikin for Acoustic Research (KEMAR). Each HRTF was represented as the impulse response of a 100-tap finite impulse response (FIR) filter at a sampling rate of 24 kHz. Separate filters were created for the left and right ears at each azimuth. Anatomical symmetry was assumed, so the right ear HRTF for azimuth A° was identical to the left ear HRTF for azimuth 360 A°. Details of the process by which the HRTFs were created are given in Chan *et al.* (2008).

The ILD-only HRTFs for each ear had the magnitude responses of the original HRTFs, with the ITD cues eliminated by using the identical phase responses for both ears, derived from the right ear HRTF for a source at 180° (i.e., back center). The ITD-only HRTFs for each ear had the phase response of the original HRTFs, with the ILD cues eliminated by using the identical magnitude response for both ears, derived from the right ear HRTF for a source at 180°.

The testing procedure was similar to that in Chan *et al.* (2008). Participants were asked to locate the stimulus (a broadband impulsive gunshot sound) presented from 1 of 12 locations, ranging from 97.5° to 262.5° (i.e., always located behind the participant) and spaced 15° apart. The locations were numbered from 1–12, with 1 located at 97.5° and 12 located at 262.5°. The participant's task was to identify the location that the stimulus originated from. Prior to testing in each condition, participants were familiarized with the real (SF) or virtual (DC) space by listening to the stimulus presented at each location once in ascending and once in descending order. For DC testing, familiarization also included a reference stimulus presented immediately prior to each stimulus, located at 90° and at 270° when familiarizing in the ascending and descending orders, respectively. After familiarization, participants were presented with a practice block followed by the test. Neither the practice block nor the test contained reference stimuli. For practice and test (but not for familiarization), roving was applied to assure that the task was testing participant's use of ITD and ILD cues rather than their ability to discriminate loudness. For SF testing, the presentation level for each stimulus was roved by a randomly selected offset in the range of ±5 dB. For DC testing, the presentation level was roved by using the overall magnitude of a randomly selected location while maintaining the ILD of the target location. To achieve comparable performance between SF and DC testing, the target was presented once at each location for SF testing and twice for DC testing prior to the participant identifying the location (Chan *et al.*, 2008). The number of stimuli presented during the test varied based on the participant's localization performance (Chan, *et al.*, 2008). All participants completed one block of 24 stimuli. Participants who, at minimum, were able to distinguish adjacent groups of three locations with an accuracy of at least 75% for the first block were presented with another block of 24 stimuli.

3. Results

The first goal was to verify that the HRTFs accurately simulated the SF. To compare localization performance between SF and DC testing, root-mean-square (RMS) error scores were calculated for each participant in each condition (Fig. 1). RMS error scores were used because, in addition to being sensitive to information across all locations, this type of analysis is also sensitive to a wide range of changes in the target-response relationship. A paired *t*-test indicated no significant difference between SF and DC testing ($t(9) = -0.8571$, $p = 0.4$), with the two testing modes resulting in very similar mean performance (SF = 26.6°, DC = 24.9°).

The second goal was to determine if the reduced availability of ITD cues for the CI user reported by Seeber and Fastl (2008) was also found for this larger set of participants. Figure 2 shows RMS error scores per listener in the three DC conditions: ITD+ILD (left panel), ITD-only (middle panel), and ILD-only (right panel). A one-way repeated measures analysis of variance revealed a significant main effect of condition [$F(2, 10) = 24.7$, $p < 0.001$]. Paired *t*-tests indicated that performance was significantly worse in the ITD-only condition (mean = 57.7°) than in either of

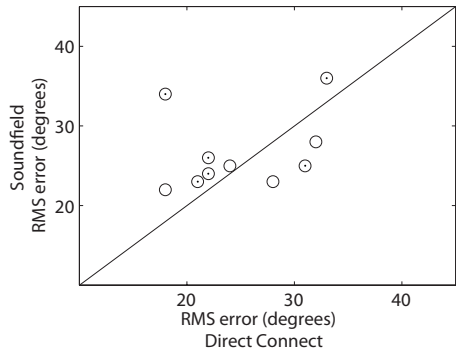


Fig. 1. Scatter plot of RMS error scores obtained when testing in the soundfield and with the ITD+ILD condition of the direct connect testing. Lower scores indicate better performance. Each circle represents one participant. The diagonal line indicates equivalence between the two testing conditions. Dotted circles demarcate participants who also completed the ITD-only and ILD-only conditions.

the other conditions (ITD+ILD mean=24.5° [$t(5)=-5.0$, $p<0.01$]; ILD-only mean=24.5° [$t(5)=-5.4$, $p<0.01$]). Every participant demonstrated this pattern. There was no significant difference in performance between the ITD+ILD and the ILD-only conditions [$t(5)=0$, $p=1$], both having the same mean RMS error. These results indicated that ILD cues played a primary role in the participants' localization performance.

The third goal was to determine whether bilateral CI users can make use of ITD cues, albeit to a lesser degree than ILD cues. Grantham *et al.* (2007) evaluated whether CI users were performing better than chance by using a computer simulation approach, whereby target lists were randomly generated and paired with randomly generated response sets. This process was repeated to derive a 95% confidence interval for chance performance.

The method of Grantham *et al.* (2007) does not take into account the effects of response bias, e.g., that a given subject might use only a subset of all possible responses. When random guessing is combined with a response bias, the obtained scores can deviate significantly from those resulting from unbiased random guessing. Consequently, confidence intervals were calculated for random guessing that corrected for the response bias for each participant and condition. This was done by randomly sampling with replacement from a given participant's response set for a particular condition to create a generated response set that incorporated the participant's response bias. That generated response set was combined with a randomly generated target list. Ten thousand target-response sets were created per participant and condition following this procedure, and RMS error was calculated for each of these target-response sets.

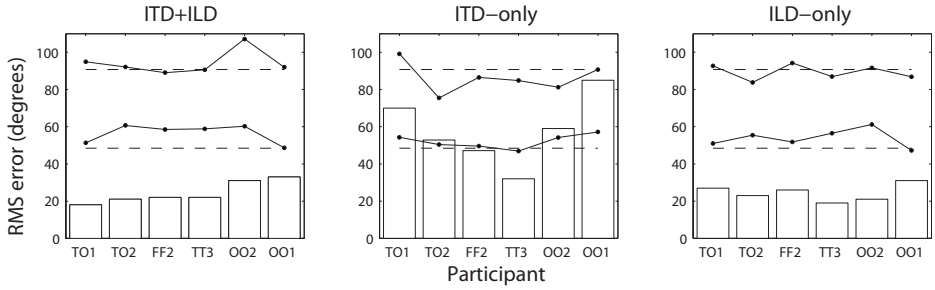


Fig. 2. Bar graphs of the RMS error scores for each participant for the three direct connect testing conditions. Lower scores indicate better performance. The dots connected with solid lines indicate the upper and lower bounds for the 99.2% confidence interval (α corrected for multiple comparisons) for random guessing corrected for the response bias of each participant. For reference, the upper and lower bounds for the 99.2% confidence interval for unbiased chance is shown as dashed lines, calculated as in Grantham *et al.* (2007).

To control for familywise error, a Bonferroni correction was used, setting α to 0.0083. The resulting confidence interval was defined as the range that encompassed 99.2% of the RMS error scores resulting from this simulation (marked by dots connected with solid lines in Fig. 2). Any participant whose RMS error score fell below the lower line of the confidence interval shown in Fig. 2 demonstrated a significant ability to use ITD cues. For comparison, chance, defined using the method of Grantham *et al.* (2007) but with α=0.0083, is also presented (confidence interval based on 10,000 randomly generated target-response sets: 48.4° and 90.7°; dashed lines in Fig. 2).

Comparison of the ITD-only scores to the biased random guessing confidence intervals (Fig. 2, middle panel) revealed that two participants performed significantly better than chance when response biases were taken into account (FF2 and TT3), indicating that some of the CI users had a significant ability to localize using interaural temporal cues only.

4. Discussion

This study examined the relative contribution of ITD and ILD cues for bilateral CI users' sound localization. The effects of these cues were independently measured using generic HRTFs designed for each type of CI processor. These HRTFs were verified by comparing each participant's RMS error scores for DC (using the HRTFs) and SF testing. These comparisons revealed similar means and no significant difference between DC and SF testing (Fig. 1).

Comparison of the RMS error scores across the three DC conditions revealed identical mean performance in the ITD+ILD and the ILD-only conditions, and significantly poorer performance in the ITD-only condition (Fig. 2). This pattern of poorer performance in the ITD-only condition was found for every participant and is consistent with the findings using indirect measures (e.g., van Hoesel and Tyler, 2003; Schoen *et al.*, 2005; Grantham *et al.*, 2007; Poon *et al.*, 2009). Despite ILD cues having a primary role, two participants were able to localize significantly better than chance in the ITD-only condition, even when response biases were taken into account (Fig. 2, middle panel).

A closer examination of Fig. 2 suggests that there was a complex interaction between ITD and ILD cues. For example, participants TO1 and OO1 were the two worst performers in both the ITD-only and the ILD-only conditions, yet TO1 was the best performer and OO1 was the worst performer for the ITD+ILD condition.

In summary, this study directly compared the contributions of ITD and ILD cues for bilateral CI users' localization, using the same measure, localization in terms of RMS error, to directly compare the two. The results demonstrated a strong reliance on ILD cues, although the interaction between ITD and ILD cues was complex. The results also indicated significant localization abilities when only using ITD cues for two participants, indicating that some bilateral CI users are sensitive to ITD cues, although further studies are needed to determine how widespread such sensitivity is across CI users. This pattern of results is quite different than that seen with individuals with unimpaired acoustic hearing, for whom ITD cues play a larger role than ILD cues for broadband sounds (e.g., Macpherson and Middlebrooks, 2002).

The results also suggest that improved temporal information may help some CI users. In current clinical practice, a variety of factors reduce the availability of temporal cues including the use of processing strategies that reduce temporal information, poor place pitch matching across interaural electrode pairs, and a lack of synchronization between the two processors. Given the evidence provided above that at least some CI users are sensitive to ITD cues, modifications to current practices that result in a more faithful transmission of temporal information may significantly improve bilateral performance.

Acknowledgments

We thank our participants for their time and effort. This work was supported by the NIH under Grant No. R44DC005759.

References and links

Chan, J. C., Freed, D. J., Vermiglio, A. J., and Soli, S. D. (**2008**). "Evaluation of binaural functions in bilateral cochlear implant users," Int. J. Audiol. **47**, 296–310.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (**2007**). "Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants," Ear Hear. **28**, 524–541.
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.
Poon, B. B., Eddington, D. K., Noel, V., and Colburn, H. S. (**2009**). "Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing," J. Acoust. Soc. Am. **126**, 806–815.
Schoen, F., Mueller, J., Helms, J., and Nopp, P. (**2005**). "Sound localization and sensitivity to interaural cues in bilateral users of the Med-El Combi 40/40+cochlear implant system," Otol. Neurotol. **26**, 429–437.
Seeber, B. U., and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am. **123**, 1030–1042.
van Hoesel, R. J. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.
van Hoesel, R. J., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Retrieving the exact Green's function by wavefield crosscorrelation

Yingcai Zheng
Department of Earth and Planetary Sciences, University of California, Santa Cruz, California 95064
yzheng@pmc.ucsc.edu

Abstract: Recent development on the Green's function retrieval by wavefield crosscorrelation has substantially advanced the physical research in a multidisciplinary and unprecedented fashion. However, the underlying assumption of the theory that the sources are in the far-field limits the technology to extracting only the high-frequency part of the Green's function in an open system. This critical approximation can be eliminated using the exact boundary integral equation method. A scheme involving the crosscorrelation kernel is proposed to recover the exact Green's function including all-frequency content. Symmetric difference kernels are analytically constructed for sources on a plane or on a circle and can be reduced to the known Dirac delta kernel under the far-field approximation.

PACS numbers: 43.20.Fn, 43.20.Bi [AN]
Date Received: October 19, 2009 **Date Accepted:** December 17, 2009

Crosscorrelating the diffuse wavefield recordings generated by random noise sources at two locations reproduces the Green's function that would be observed at one location as if a source were at the other location (e.g., Lobkis and Weaver, 2001; Weaver and Lobkis, 2001). No technology has been comparable to the Green's function retrieval in recent times that has produced far-reaching consequences across so many physical disciplines such as in ultrasonics (Lobkis and Weaver, 2001), acoustics (Derode *et al.*, 2003), ocean acoustics (Roux and Fink, 2003), terrestrial seismology (Aki, 1957; Claerbout, 1968; Campillo and Paul, 2003; Calvert *et al.*, 2004; Schuster *et al.*, 2004; Bakulin and Calvert, 2006; Schuster and Zhou, 2006), helioseismology (Duvall *et al.*, 1993; Rickett and Claerbout, 1999), and medical diagnostics (Sabra *et al.*, 2007). The emergence of the Green's function by wavefield crosscorrelation has been theorized using the normal modes based on the energy equipartition principle in which the wavefield is diffuse and waves propagate in all directions for both closed (Lobkis and Weaver, 2001; Tanimoto, 2008) and open media (Weaver and Lobkis, 2004). When the inhomogeneous medium is not illuminated by random noise sources and the diffusivity of the wavefield is not achieved, theoretical approaches including the stationary phase method (e.g., Schubert, 2001; Snieder, 2004) and the representation theorem (Wapenaar, 2004) were used to demonstrate the applicability of the technology. However, the theories, in particular, for the open system, invoked an assumption that the illuminating sources are in the far-field regime may be violated in important applications. The consequence of this critical approximation is that the retrieved Green's function contains incorrect amplitude and it consists of only high-frequency propagating waves and no evanescent waves which are crucial in applications such as the super-resolution diffraction tomography in medical imaging or mining activity (Lehman, 2002). In this letter, we show that the far-field approximation can be removed by the exact boundary integral equation method and the exact Green's function with full-frequency bandwidth can be retrieved, including all types of scattering effects. Our exact method leads to construction of a crosscorrelation kernel and because this kernel is same for both deterministic and noise sources, we consider the former case in the sequel.

The Green's function retrieval can be understood by a surface integral (Fig. 1) in the frequency ω domain using the Rayleigh's reciprocity of the correlational type (e.g., Wapenaar and Fokkema, 2006; Shuster, 2009):

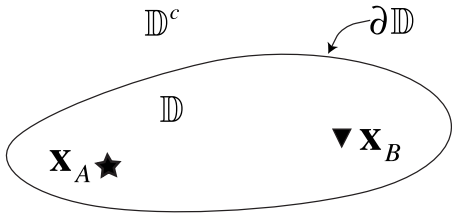


Fig. 1. Heterogeneous medium bounded by a surface $\partial\mathbb{D}$.

$$2i\ \mathrm{Im}\ G(\mathbf{x}_B|\mathbf{x}_A;\omega) = \oiint_{\partial\mathbb{D}} \left[G(\mathbf{x}'|\mathbf{x}_A;\omega)\frac{\partial\bar{G}(\mathbf{x}'|\mathbf{x}_B;\omega)}{\partial n'} - \bar{G}(\mathbf{x}'|\mathbf{x}_B;\omega)\frac{\partial G(\mathbf{x}'|\mathbf{x}_A;\omega)}{\partial n'} \right] d^2\mathbf{x}' \tag{1}$$

in which $G(\mathbf{x}'|\mathbf{x}_A;\omega)$ and $G(\mathbf{x}'|\mathbf{x}_B;\omega)$ are the Green's functions for the heterogeneous medium in $\mathbb{D}$ which satisfy differential equations

$$\nabla^2 G(\mathbf{x}'|\mathbf{x}_{A,B};\omega) + k^2(\mathbf{x}')G(\mathbf{x}'|\mathbf{x}_{A,B};\omega) = -\delta(\mathbf{x}'-\mathbf{x}_{A,B}), \quad \mathbf{x}',\mathbf{x}_{A,B} \in \mathbb{D}\cup\partial\mathbb{D},$$

and the overbar denotes the complex conjugate and $\partial/\partial n'$ is the directional derivative along the outward normal of the boundary $\partial\mathbb{D}$ with respect to position $\mathbf{x}' \in \partial\mathbb{D}$. For simplicity, we neglect the source time function in the formulation. The right hand side of Eq. (1) can be interpreted as focusing/backpropagating the wavefield on $\partial\mathbb{D}$ generated by a source at $\mathbf{x}_A$ to the observation point $\mathbf{x}_B$. Based on a physical argument using the time reversal acoustics (Derode *et al.*, 2003), the left hand side of Eq. (1) corresponds to $G(\mathbf{x}_B|\mathbf{x}_A;t)-G(\mathbf{x}_B|\mathbf{x}_A;-t)$ in the time t domain, which consists of the causal and the anticausal Green's functions. Let $p_A(\mathbf{x}')=G(\mathbf{x}'|\mathbf{x}_A;\omega)$ and $q_A(\mathbf{x}')=\partial G(\mathbf{x}'|\mathbf{x}_A;\omega)/\partial n'$ be the wavefields on $\partial\mathbb{D}$ produced by a source at $\mathbf{x}_A$. Likewise, $p_B(\mathbf{x}')=G(\mathbf{x}'|\mathbf{x}_B;\omega)$ and $q_B(\mathbf{x}')=\partial G(\mathbf{x}'|\mathbf{x}_B;\omega)/\partial n'$ are due to the source at $\mathbf{x}_B$. Equation (1) can be written in a compact form

$$2i\ \mathrm{Im}\ G(\mathbf{x}_B|\mathbf{x}_A;\omega) = \langle p_A, q_B\rangle - \langle q_A, p_B\rangle \tag{2}$$

in which $\langle f,g\rangle=\int_{\partial\mathbb{D}} f(\mathbf{x}')\bar{g}(\mathbf{x}')d^2\mathbf{x}'$ is the inner product defined on the boundary. In previous studies (e.g., Wapenaar *et al.*, 2005; Wapenaar and Fokkema, 2006; Snieder *et al.*, 2007), the integral was simplified in an open system using the far-field approximation for a spherical $\partial\mathbb{D}$ with a large radius such that $q_{A,B}(\mathbf{x}')\approx i\omega c^{-1}(\mathbf{x}')p_{A,B}(\mathbf{x}')$, $\mathbf{x}' \in \partial\mathbb{D}$. If the wave propagation speed $c(\mathbf{x}')$ is constant, we obtain

$$2i\ \mathrm{Im}\ G(\mathbf{x}_B|\mathbf{x}_A;\omega) \approx -2ik\langle p_A, p_B\rangle, \quad k=\omega/c, \tag{3}$$

which is referred to as Green's function retrieval from far-field correlations. By the source-receiver reciprocity in the Green's function, $p_A(\mathbf{x}')$ and $p_B(\mathbf{x}')$ can be interpreted as the wavefields recorded at $\mathbf{x}_A$ and $\mathbf{x}_B$, respectively, produced by a source at $\mathbf{x}' \in \partial\mathbb{D}$. Each source contributes equally to the Green's function retrieval on the left hand side. Next, we show that one can still relate the integral (1) to the crosscorrelation of the wavefield as in Eq. (3) even when $\partial\mathbb{D}$ is not in the far-field regime.

It is well known in the studies of the Huygens principle using single-layer and double-layer potential representations that the surface values p_A and q_A (or p_B and q_B) are generally not independent (e.g., Baker and Copson, 1950; Kupradze, 1963; Colton and Kress, 1983). To see this, we consider the scattering problem due to a source at $\mathbf{x}_A \in \mathbb{D}$ (Fig. 1). In the exterior domain $\mathbb{D}^c$, we apply the boundary integral equation method (e.g., Kleinman and Roach, 1974; Sanchez-Sesma and Campillo, 1991; Rjasanow and Steinbach, 2007):

$$\sigma(\mathbf{x}'')p_A(\mathbf{x}'') = \oint\!\!\!\!\!\!\int_{\partial\mathbb{D}} \left[p_A(\mathbf{x}')\frac{\partial G(\mathbf{x}'|\mathbf{x}'';\omega)}{\partial n'} - q_A(\mathbf{x}')G(\mathbf{x}'|\mathbf{x}'';\omega) \right] d^2\mathbf{x}', \quad \mathbf{x}',\mathbf{x}'' \in \partial\mathbb{D} \tag{4}$$

in which the continuity of p_A and q_A across the boundary is imposed and the integral is in the sense of Cauchy principal value. If the boundary is not smooth at $\mathbf{x}'' \in \partial\mathbb{D}$ (e.g., a vertex), $\sigma(\mathbf{x}'')$ is related to the solid angle spanned by the vertex locally and $\sigma(\mathbf{x}'')=0.5$ for a smooth boundary point at $\mathbf{x}''$. Equation (4) can be written in the integral operator form

$$Sq_A(\mathbf{x}'') = \left(K - \frac{1}{2}I \right) p_A(\mathbf{x}''), \quad \mathbf{x}'' \in \partial\mathbb{D} \tag{5}$$

(e.g., Martin, 2006). The exterior Green's function satisfies the differential equation in the exterior domain

$$\nabla'^2 G(\mathbf{x}';\mathbf{x}'') + k^2(\mathbf{x}')G(\mathbf{x}';\mathbf{x}'') = -\delta(\mathbf{x}' - \mathbf{x}''), \quad \mathbf{x}',\mathbf{x}'' \in \mathbb{D}^c \cup \partial\mathbb{D}$$

and the Sommerfeld radiation condition at the infinity

$$\lim_{r\to\infty} r\left(\frac{\partial}{\partial r} - ik \right) G(\mathbf{x}';\mathbf{x}'') = 0, \quad r = |\mathbf{x}' - \mathbf{x}''|$$

if the exterior medium is unbounded. The null space of the operator S is not empty if ω is one of the eigenfrequencies $\{\omega_k, k \in \mathbb{N}\}$ of the interior problem with the homogeneous Dirichlet boundary condition (Colton and Kress, 1983). Under this case, the inverse S^{-1} of S does not exist and q_A cannot be determined uniquely from p_A. Because the backpropagation integral is a general expression for all ω, we can restrict our discussion to the case $\omega \neq \omega_k$. However, if the exterior domain is unbounded and the Sommerfeld radiation condition is ensured at the infinity, the boundary value q_A is uniquely determined by p_A (Baker and Copson, 1950). Once the inverse S^{-1} of S exists, we have

$$q_A = S^{-1}\left(K - \frac{1}{2}I \right) p_A = Cp_A \tag{6}$$

in which I is the identity operator. The same reasoning can be applied to a source at $\mathbf{x}_B$:

$$q_B = Cp_B. \tag{7}$$

Substituting Eqs. (6) and (7) into Eq. (2), we obtain

$$2i \text{ Im } G(\mathbf{x}_B|\mathbf{x}_A;\omega) = \langle p_A, Cp_B\rangle - \langle Cp_A, p_B\rangle = \langle p_A, (C - C^*)p_B\rangle = \langle p_A, Wp_B\rangle \tag{8}$$

in which C^* is the adjoint of C. If we discretize the boundary, C will be a square matrix and C^* is the Hermitian transpose of C. The invertibility of the matrix S can be studied readily from its eigenvalues. We take the interpretation that $p_A(\mathbf{x}')$ and $p_B(\mathbf{x}')$ are wavefields at $\mathbf{x}_A$ and $\mathbf{x}_B$, respectively, due to a boundary source at $\mathbf{x}'$. Clearly, if $W=\{w_{ij}\}$ is a diagonal matrix, then the backpropagation is a weighted correlation, $\Sigma_i w_{ii} p_A^i \bar{q}_B^i$ in which i is the index of the source. However, if W is a full matrix, the backpropagation integral becomes $\Sigma_i \Sigma_j w_{ij} p_A^i \bar{q}_B^j$ and i and j are the source indices. In this case, the wavefield p_A^i at $\mathbf{x}_A$ due to the ith source on the boundary is crosscorrelated with the wavefield p_B^j at $\mathbf{x}_B$ due to the jth source, which is in sharp contrast with the previous thought that only wavefields at two locations from the same source are crosscorrelated. In general, the matrix W depends on the exterior medium and the boundary geometry. Note that if the exterior medium is unbounded and homogeneous, W only depends on the geometry of the boundary ∂D and the wavenumber k in the exterior medium.

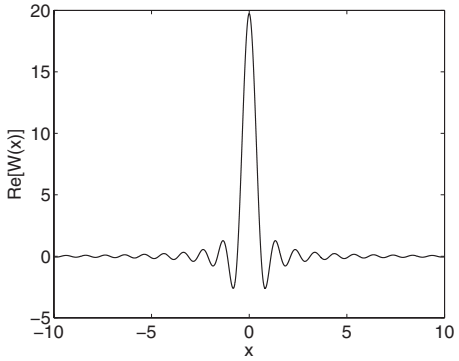


Fig. 2. Crosscorrelation kernel for the plane surface case. The wavelength $\lambda = 1$.

We derive the crosscorrelation kernels for two important boundaries frequently appearing in applications: an infinite plane boundary and a cylindrical boundary. Both problems are formulated in the two-dimensional space and both kernels are symmetric difference kernels whose actions on a function represent convolution operations.

Example 1. In the two-dimensional space, let the boundary $\partial\mathbb{D}$ be an infinitely large plane. In this case, $K=0$. The integral operator kernel is

$$S = \frac{i}{4} H_0(k|x|) \tag{9}$$

in which H_0 is the Green's function in the two-dimensional space. Because the kernel S is a symmetric difference kernel whose action on a function is a convolution, the Fourier transform in the distributional sense can be applied to compute its inverse and the crosscorrelation kernel reads

$$W(x) = ikJ_1(k|x|)/|x|, \tag{10}$$

where J_1 is the first order Bessel function of the first kind. Figure 2 plots one such a kernel.

Example 2. If the surface is a cylinder with radius R and the exterior medium is unbounded and homogeneous and has a wavenumber k, the integral equation in the exterior domain reads

$$\frac{1}{2}u(\theta) = -\frac{i}{4}\int_0^{2\pi} q(\theta')H_0\left(2kR\left|\sin\frac{\theta-\theta'}{2}\right|\right)Rd\theta' - \frac{ik}{4}\int_0^{2\pi} u(\theta')H_1\left(2kR\left|\sin\frac{\theta-\theta'}{2}\right|\right) \times \left|\sin\left(\frac{\theta-\theta'}{2}\right)\right| Rd\theta' \tag{11}$$

in which polar angles θ and θ' denote positions on the boundary. All the functions in the above integral are 2π-periodic. Therefore, we can use the Fourier series to obtain the crosscorrelation kernel in the discrete angular wavenumber domain

$$W_n = \frac{4i}{\pi R}|H_n(kR)|^{-2}, \quad n = 0, \pm 1, \pm 2, \ldots \tag{12}$$

in which W_n is the coefficient of the Fourier series for the expansion function $e^{in\theta}$ and H_n the first-kind Hankel function of order n. For the far-field case, $kR \gg 1$, our crosscorrelation kernel reduces to the known result, $W(\theta) = 2ik\delta(\theta)$, where $\delta(\theta)$ is the Dirac delta function. Hence we have rigorously justified the validity of using the far-field approximation in the Green's function retrieval in Eq. (3). Let $R=1$ and we compute the crosscorrelation kernels for several wavenumbers (Fig. 3). At the high frequency, the crosscorrelation kernel for the cylindrical boundary resembles that of the plane boundary case [Figs. 3(c) and 2], which indicates that if the radius of curvature is large compared to the wavelength, the kernel of the plane boundary may be applied

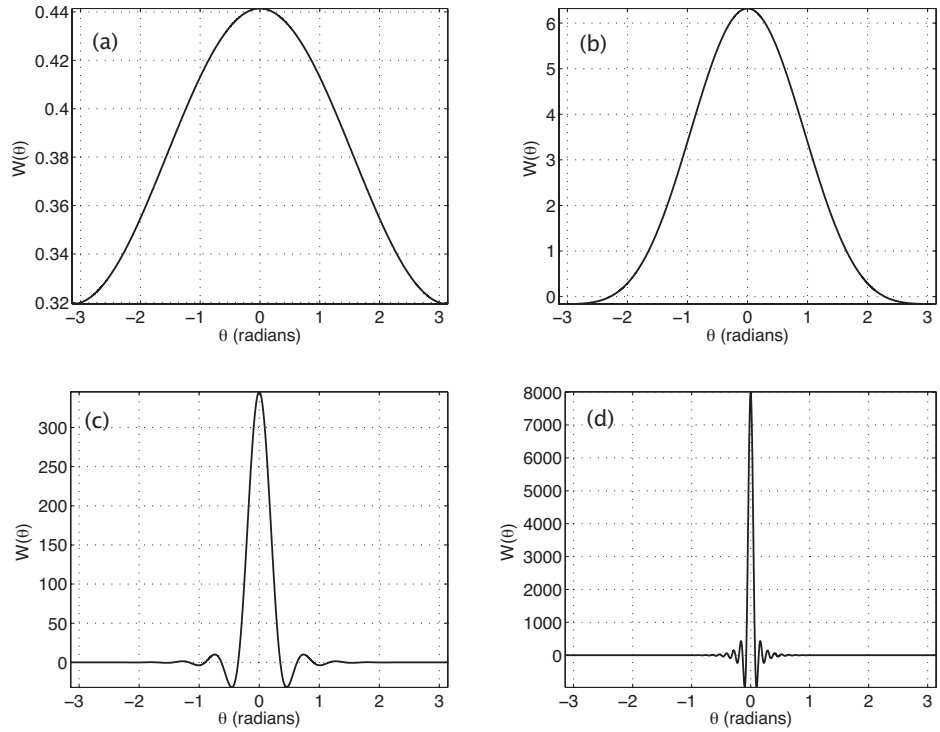


Fig. 3. Crosscorrelation kernels for a cylinder boundary with radius $R=1$ for different wavenumbers (a) $k=0.1$, (b) $k=1$, (c) $k=10$, and (d) $k=50$. The kernels are obtained by inverse Fourier transform of Im W_n.

locally. For both the plane and the cylindrical boundaries, the kernels approach the Dirac delta distribution if we increase the frequency ω. Numerical examples for the cylindrical boundary corroborated our theoretical prediction that the accuracy of the retrieved Green's function increases for large kR if we use the delta kernel (Wapenaar *et al.*, 2005).

In this letter, we have pointed out that the previous theory with the far-field approximation on the Green's function retrieval yielded incorrect amplitude and prevented its use from extracting the low-frequency content of the Green's function in an open system. We proposed a method based on the integral equation method to reconstruct the exact full Green's function including all wave phenomena if source locations are known such as in the seismic interferometry using controlled sources. If deterministic sources were used, we have shown that in order to exactly retrieve the Green's function we need to correlate wavefields not only between receivers for the same source but also between different sources. Finally, we remark that for a great number of applications, the Green's function retrieval technique is being applied to diffuse wavefields generated by noise sources and as such the crosscorrelation kernel should still be used if the source locations can be estimated. For a boundary with general geometric shape on which the sources are excited, the boundary element method can be used to solve for the crosscorrelation kernel.

I am grateful to Associate Editor Professor Andrew Norris and an anonymous reviewer for their constructive comments. I also thank Ru-shan Wu, Xiao-bi Xie, Yaofeng He, Rui Yan, and Qiaoer Zhou for verifying the idea and checking the derivation. The discussion with Professor Haruo Sato on the far-field approximation was illuminating. I thank Professor Harold Widom for the help on the integral equation with Hankel kernel and its inverse. I thank Dr. Brian Schlottmann and Dr. Heather Savage for reading the manuscript and their useful comments. This work is supported by the NSF Grant No. EAR-0838359 and by the Wavelet Transform on Propagation and Imaging (WTOPI) Consortium/UCSC. Facility support from the W. Keck Foundation is appreciated.

References and links

Aki, K. (**1957**). "Space and time spectra of stationary stochastic waves with special reference to microtremors," Bull. Earthquake Res. Inst., Univ. Tokyo **5**, 415–456.

Baker, B. B., and Copson, E. T. (**1950**). *The Mathematical Theory of Huygens' Principle*, 2nd ed. (Clarendon, Oxford).

Bakulin, A., and Calvert, R. (**2006**). "The virtual source method: Theory and case study," Geophysics **71**, SI139–SI150.

Calvert, R. W., Bakulin, A., and Jones, T. C. (**2004**). "Virtual sources, a new way to remove overburden problems," in 66th Meeting EAGE Expanded Abstracts, p. 234.

Campillo, M., and Paul, A. (**2003**). "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549.

Claerbout, J. F. (**1968**). "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269.

Colton, D., and Kress, R. (**1983**). *Integral Equation Methods in Scattering Theory* (Wiley, New York).

Derode, A., Larose, E., Campillo, M., and Fink, M. (**2003**). "How to estimate the Green's function of a heterogeneous medium between two passive sensors? Application to acoustic waves," Appl. Phys. Lett. **83**, 3054–3056.

Duvall, T. L., Jefferies, S. M., Harvey, J. W., and Pomerantz, M. A. (**1993**). "Time distance helioseismology," Nature (London) **362**, 430–432.

Kleinman, R. E., and Roach, G. F. (**1974**). "Boundary integral equations for the three-dimensional Helmholtz equation," SIAM Rev. **16**, 214–236.

Kupradze, V. D. (**1963**). *Dynamical Problems in Elasticity* (North-Holland, Amsterdam).

Lehman, S. K. (**2002**). "Superresolution planar diffraction tomography through evanescent fields," Int. J. Imaging Syst. Technol. **12**, 16–26.

Lobkis, O. I., and Weaver, R. L. (**2001**). "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017.

Martin, P. A. (**2006**). *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles* (Cambridge University Press, New York).

Rickett, J., and Claerbout, J. (**1999**). "Acoustic daylight imaging via spectral factorization; helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960.

Rjasanow, S., and Steinbach, O. (**2007**). *The Fast Solution of Boundary Integral Equations* (Springer Science+Business Media, LLC, New York).

Roux, P., and Fink, M. (**2003**). "Green's function estimation using secondary sources in a shallow water environment," J. Acoust. Soc. Am. **113**, 1406–1416.

Sabra, K. G., Conti, S., Roux, P., and Kuperman, W. A. (**2007**). "Passive in vivo elastography from skeletal muscle noise," Appl. Phys. Lett. **90**, 194101–1–194101–3.

Sanchez-Sesma, F. J., and Campillo, M. (**1991**). "Diffraction of P, SV, and Rayleigh waves by topographic features: A boundary integral formulation," Bull. Seismol. Soc. Am. **81**, 2234–2253.

Schubert, G. (**2001**). "Seismic interferometric/daylight imaging: Tutorial," in 63rd Annual Conference, EAGE Extended Abstracts.

Schuster, G. T., Yu, J., Sheng, J., and Rickett, J. (**2004**). "Interferometric/daylight seismic imaging," Geophys. J. Int. **157**, 838–852.

Schuster, G. T., and Zhou, M. (**2006**). "A theoretical overview of model-based and correlation-based redatuming methods," Geophysics **71**, SI103–SI110.

Shuster, G. T. (**2009**). *Seismic Interferometry* (Cambridge University Press, Cambridge,.

Snieder, R. (**2004**). "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610.

Snieder, R., Wapenaar, K., and Wegler, U. (**2007**). "Unified Green's function retrieval by cross-correlation; connection with energy principles," Phys. Rev. E **75**, 036103–1–036103–14.

Tanimoto, T. (**2008**). "Normal-mode solution for the seismic noise cross-correlation method," Geophys. J. Int. **175**, 1169–1175.

Wapenaar, K. (**2004**). "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301.

Wapenaar, K., and Fokkema, J. (**2006**). "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46.

Wapenaar, K., Fokkema, J., and Snieder, R. (**2005**). "Retrieving the Green's function in an open system by cross correlation: A comparison of approaches," J. Acoust. Soc. Am. **118**, 2783–2786.

Weaver, R. L., and Lobkis, O. I. (**2001**). "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301–1–134301–4.

Weaver, R. L., and Lobkis, O. I. (**2004**). "Diffuse fields in open systems and the emergence of the Green's function (L)," J. Acoust. Soc. Am. **116**, 2731–2734.

On the subtraction method for *in-situ* reflection and diffusion coefficient measurements

Philip Robinson and Ning Xiang
Graduate Program in Architectural Acoustics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180
robinp@rpi.edu, xiangn@rpi.edu

Abstract: The subtraction method is a technique critical to several important acoustic measurements. It involves subtracting a reference measurement including only direct sound from one with direct sound and a reflection, to isolate the reflection. The process is very sensitive to environmental conditions, such as changes in temperature, air movement, and microphone positioning. These variations cause small time differences between the reference and reflection measurements, which prevent complete subtraction of the direct sound; the residual direct sound then pollutes analysis of the isolated reflection. This work evaluates methods to compensate for differences to achieve minimal interference from the residual direct sound.

PACS numbers: 43.58.Bh, 43.58.Vb, 43.55.Ev [MS]
Date Received: November 4, 2009 **Date Accepted:** January 4, 2010

1. Introduction

The subtraction method is an important step in several acoustic measurements that are important to acoustic practitioners in the field. The first is the *in-situ* absorption/reflection coefficient measurement. This method was suggested by Yuzawa[1] as early as 1975 and further developed by Mommertz.[2] Nocke[3] and Garai[4] also utilized this technique. ISO Standard 13472 (Ref. 5) for the measurement of the absorption of roadway surfaces incorporates the subtraction technique and has been applied to other surfaces such as grass and turf.[6] The measurement involves taking an impulse response in the free field, and another with the microphone close to the surface under test. The free field measurement is subtracted from the test measurement to isolate the reflection, which is then compared to the free field direct sound to deduce the reflection coefficient from the surface. This measurement technique is of high practical significance because it allows testing of installed materials without the need to remove material to a laboratory for reverberation chamber[7] or impudence tube[8] measurements.

The second measurement that requires this technique is the measurement of diffusion coefficients.[9] Diffusion coefficients are measured by placing a semicircular or hemispherical array of microphones around the sample to be measured, a source outside of the microphone array, and taking impulse responses for each microphone. The polar response of the panel is then distilled into a diffusion coefficient using an auto-correlation function as in Eq. (2). In attaining the polar response from an architectural wall panel for many source incidences, it is often necessary to place the source at grazing incidence to the panel with the microphone opposite. Figure 1 (right) illustrates this condition. This results in the direct sound and panel reflection arriving at the microphone in close succession. Since only the reflection is of interest, a second measurement is taken without the panel present to attain a reference that can be subtracted from the measurement to isolate the panel reflection. This measurement is a simple and practical method to measure surfaces such as those described by Olson and Bradley,[10] Dadiotis *et al.*,[11] and D'Antonio,[12] among others. Both of these measurements require taking either two measurements with the same microphone or taking one measurement with two different microphones. In either case, there will be slight differences between the measurements. These differences arise from differences in the response of the microphones, no matter how closely matched

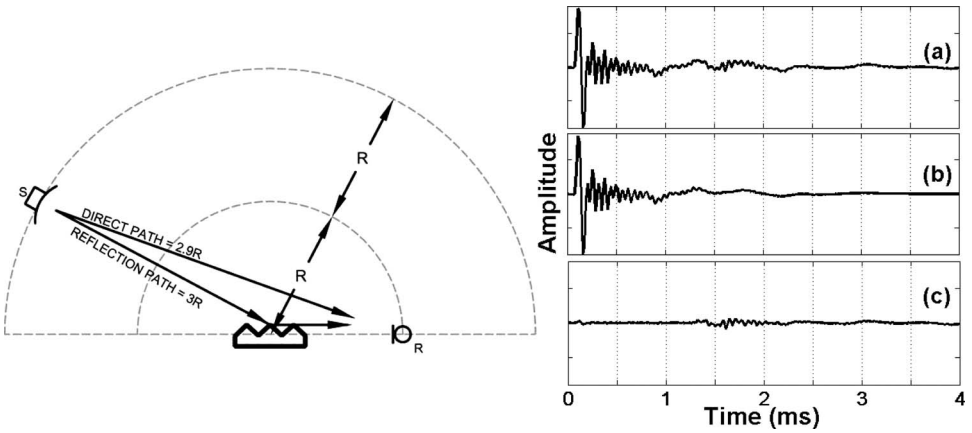


Fig. 1. Diffusion coefficient measurement setup that requires the subtraction method to isolate the reflection and the resulting impulse responses: (a) Impulse response of reflections including the direct sound impulse. (b) Reference impulse response. (c) Subtraction results.

they are, temperature changes and air movement between measurements, and variations in the response of the measurement equipment due to temperature or electrical deviations. This work examines methods to compensate for these small changes.

2. Experiment description

This experiment uses impulse responses obtained from diffusion coefficient measurements. A measurement as shown in Fig. 1 (left) without the diffuser panel present yields the reference measurement; one with the panel present provides the reflection measurement. Subtracting the former from the latter eliminates the direct sound and leaves only the reflection. This paper examines a variety of methods to correct for minor time misalignments between the two signals to attain the optimal result. The subtraction results from each method are then used to calculate the diffusion coefficient,[13] and variations are analyzed. Figure 1 illustrates the measured impulse responses and subtraction results.

3. Subtraction methods

The first scheme examined is the direct subtraction method. This method assumes that environmental conditions are constant between measurements and measurements can be repeated exactly. If this assumption is correct, direct subtraction should produce complete elimination of the direct sound and leave only the isolated reflection. If this is not effective, which is often the case, further processing is necessary.

Several steps may be taken to improve the results. All of these steps require comparing different features of the two measurements. To make this comparison more accurate it is desirable to oversample both signals so corrections of less than one sampling point can be made; after subtraction the signals are downsampled to their original sampling frequency.

The simplest correction is to calibrate the signals so that the amplitudes of the direct sound peaks are the same. This ensures that differences in magnitude do not skew the results and do not account for time misalignments. Aligning the signals in time further improves the subtraction results. This can be done in several ways. Utilizing the peaks of the direct sounds as a reference point and aligning them in time can help correct for overall shifts, but a small shift in the peak may not accurately represent the shift of the whole signal. To incorporate a larger portion of the signal into consideration, cross-correlation of a few milliseconds of the signals, centered on the direct sound peaks, may be effective. The location of the peak of the cross-correlation function will indicate how much to shift one signal in relation to the other to attain

the best alignment of the direct sounds. However, the cross-correlation function is often skewed by the presence of the reflection within the direct sound component. This is often the case when measuring the reflection coefficient of a rigid surface when the microphone is close to the surface.

4. Subtraction optimization

The optimized approach to achieve the best alignment includes all of the previous schemes, then shifts the signals across each other several points to either side of the shift indicated by the cross-correlation function. By subtracting the reference from the reflection measurement at each of these locations, the most effective result can be found. This is performed at a wide range of oversampling rates in order to have the ability to shift the signals by any fraction of a sampling point. So, for an oversampling rate of 2, the signals are shifted four points in either direction; for 3, six points in either direction; etc. This is repeated up to an oversampling rate of 20 or more, so that every fraction of shift within the four points around the peak correlation is investigated. There is some redundancy since, for example, 10/20 is the same as 1/2, but the advantage is that a step of, say, 1/19 of a point is investigated. At each shift the signals are subtracted and the residual direct sound is compared. More thorough organization of the oversampling rates could improve calculation time by removing the redundant steps. For a limited number of oversampling rates and shifts, a gridded search to find the most effective subtraction is sufficient to obtain optimized results; in more detailed investigations a genetic algorithm can be exploited to improve computation time.

5. Measuring subtraction effectiveness

The process above provides many varying subtraction results; in order to choose the best result, a metric of success is required. As the goal of the operation is to remove the direct sound, subtraction effectiveness can be measured by the percent reduction, or decibel level reduction in the direct sound from the measurement to the result. Specifically, the sum of the energy within 0.5 ms of either side of the direct sound can be compared before and after subtraction to find the effective reduction. Equation (1) defines the reduction factor R. A reduction factor equal to the peak to noise ratio of the measurement can be considered a complete subtraction, since this would leave nothing in the area of the direct sound except the background noise.

$$R = 10\ \log_{10}\left(\frac{\sum_{ds-0.5\ \mathrm{ms}}^{ds+0.5\ \mathrm{ms}} (S_{\mathrm{ref}})^2}{\sum_{ds-0.5\ \mathrm{ms}}^{ds+0.5\ \mathrm{ms}} (S_{\mathrm{result}})^2}\right), \tag{1}$$

where ds is the time of the direct sound peak, S_{ref} is the reference measurement, and S_{result} is the subtraction result.

Certainly, if direct subtraction provides suitable reduction in the direct sound component, no further processing is necessary. The level of the residual direct sound in relation to the level of the reflection should also be considered. The smaller the reflection component, the greater influence a given residual direct sound will have. This is particularly important in the *in-situ* measurement of absorptive materials and grazing source incidence diffusion coefficient measurements. In both cases the reflection is small and also within close proximity to the direct sound.

6. Results

Seventy-two sets of measurements were processed with five different subtraction schemes. The schemes, as described above, are direct subtraction, peak amplitude matching (PAM), peak amplitude and time matching (PATM), peak amplitude matching with cross-correlation alignment (PACA), and the optimized method (OM). An example of the results from two different methods is shown in Fig. 2 (left). Calculating the reduction factor R per Eq. (1) for each mea-

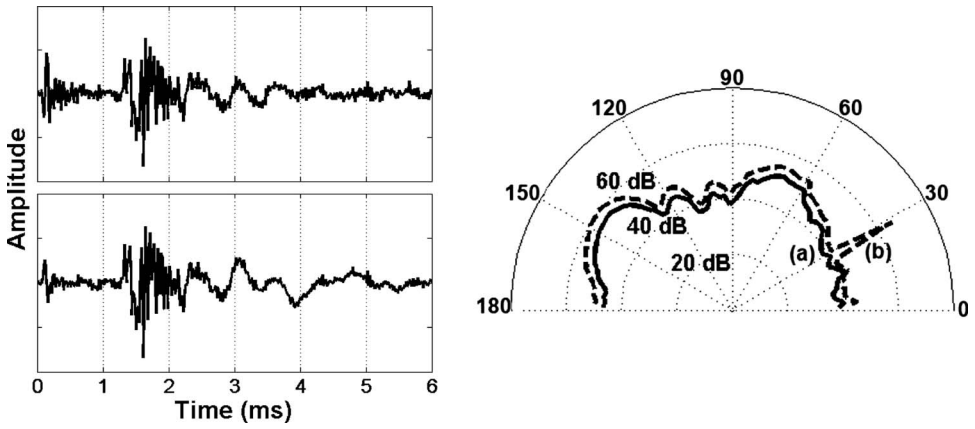


Fig. 2. (Left) A comparison of the direct subtraction result (a) and the optimized result (b). In this case, there was a 5 dB improvement in the reduction in the direct sound. (Right) A polar response with an anomalous spike from failed subtraction (b) and the corrected response (a). The diffusion coefficients are 0.19 and 0.35, respectively.

surement and subtracting the direct method reduction factor yield the improvement offered by each method. Table 1 lists the improvement results for the 72 measurement sets processed. Negative values indicate results that are worse than direct subtraction.

Peak amplitude matching provides some improvement, and since this scheme is so simple to implement, it should be used on all subtraction measurements. Note that aligning the peaks of the direct sound provides no additional improvement. This indicates that the exact alignment of the peaks does not correlate to the best time alignment of the entire direct sound. Cross-correlation, used blindly, provides generally worse results than if the signals had been subtracted directly. This is possibly due to the influence of the reflection in the correlation, which could skew the correlation peak due to the mismatch between the reflection and background noise after the direct sound. Finally, the optimization method shows a marked improvement over direct subtraction, sometimes as much as 12 dB. Where the optimization method failed, as indicated by the −1 dB improvement, was on the rare measurement where direct subtraction was extremely accurate; these measurements showed direct method R factors above all others.

7. Effect on the diffusion coefficient

Since the diffusion coefficient uses auto-correlation of many channels, per Eq. (2), it is susceptible to being skewed by one faulty channel.

Table 1. A comparison of the improvement ($R_{\text{method}}-R_{\text{direct}}$) provided by each subtraction scheme: PAM, PATM, PACA, and OM.

Method	Minimum (dB)	Mean (dB)	Maximum (dB)
PAM	−1.6	0.4	4.5
PATM	−1.6	0.4	4.5
PACA	−29.3	−2.5	4.5
OM	−1	2.6	12.1

Table 2. A comparison of the diffusion coefficients calculated from different subtraction schemes.

$\frac{1}{3}$ octave band (Hz)	Direct	PAM	PATM	PACA	Optimized
100	0.57	0.58	0.58	0.54	0.68
125	0.57	0.57	0.57	0.52	0.71
160	0.44	0.43	0.43	0.44	0.51
200	0.37	0.37	0.37	0.38	0.41
250	0.52	0.52	0.52	0.40	0.61
315	0.45	0.45	0.45	0.30	0.52
400	0.19	0.19	0.19	0.12	0.35
500	0.05	0.05	0.05	0.04	0.57
630	0.05	0.05	0.05	0.05	0.48
800	0.04	0.04	0.04	0.04	0.56
1000	0.01	0.01	0.01	0.03	0.48
1250	0.01	0.01	0.01	0.03	0.38
1600	0.01	0.01	0.01	0.03	0.36
2000	0.01	0.01	0.01	0.04	0.62
2500	0.03	0.03	0.03	0.04	0.45

$$D_\Theta = \frac{\left(\sum_{i=1}^{n} 10^{L_i/10}\right)^2 - \sum_{i=1}^{n} (10^{L_i/10})^2}{(n-1)\sum_{i=1}^{n} (10^{L_i/10})^2}, \quad (2)$$

where i is the microphone position, L_i is the level at i microphone, n is number of microphones in the array, and Θ is the source incidence.

In a semicircular array of 72 microphones, if the subtraction method fails on one or two channels, a spike can occur in the polar response. This spike is interpreted as highly directional reflection, and the diffusion coefficient will be artificially low. Figure 2 (right) shows a polar response that has a spike from failed subtraction and the corrected response calculated from optimized subtractions. The effect on the diffusion coefficient is pronounced.

Table 2 lists diffusion coefficients of a quadratic residue diffuser designed for 600–1600 Hz with the source at 60° incidence for each subtraction method. The effect of failed subtraction becomes most evident above 400 Hz where all of the non-optimized methods yield erroneous results. Even below this threshold, the optimized method, which eliminates the spikes, shows a higher diffusion coefficient indicating a more uniform polar response. The consistency of the optimized results also indicates that the method is less subject to variations in the quality of the measurement from each channel of the semicircular array.

8. Concluding remarks

The subtraction method is particularly sensitive to variations in environmental conditions between the reference measurement and the reflection measurement. Aligning the signals from these measurements in amplitude and time can eliminate the differences to provide clean subtraction results. This is a critical step in both *in-situ* reflection measurements and diffusion coefficient measurements; if the subtraction is not effective, the results from both measurements will be skewed if not entirely erroneous. This complication is amplified in diffusion coefficient measurements because the measurement gauges consistency across many channels. Shifting the signals across one another at various oversampling rates to attain the highest reduction factor is an effective method to determine the optimal subtraction result.

References and links

[1]M. Yuzawa, "A method of obtaining the oblique incident sound absorption coefficient through an on-the-spot measurement," Appl. Acoust. **8**, 27–41 (1975).
[2]E. Mommertz, "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," Appl. Acoust. **46**, 251–263 (1995).
[3]C. Nocke, "In-situ acoustic impedance measurement using a free-field transfer function method," Appl. Acoust. **59**, 253–264 (2000).
[4]M. Garai, "Measurement of the sound-absorption coefficient in situ: The reflection method using periodic pseudo-random sequences of maximum length," Appl. Acoust. **39**, 119–139 (1993).
[5]ISO 13472-1:2002, "Acoustics—Measurement of sound absorption properties of road surfaces in situ—Part 1: Extended surface method" (International Organization for Standardization, Geneva).
[6]N. Londhe, M. D. Rao, and J. R. Blough, "Application of the ISO 13472-1 in situ technique for measuring the acoustic absorption coefficient of grass and artificial turf surfaces," Appl. Acoust. **70**, 129–141 (2009).
[7]ISO 354, "Acoustics measurement of sound absorption in a reverberation room" (International Organization for Standardization, Geneva).
[8]ISO 10534-2:1998, "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method" (International Organization for Standardization, Geneva).
[9]T. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application*, 2nd ed. (Taylor & Francis, London, 2009).
[10]D. R. Olson and D. T. Bradley, "Fractal surfaces: Generation and acoustic scattering prediction," J. Acoust. Soc. Am. **125**, 2613–2613 (2009).
[11]K. Dadiotis, J. A. S. Angus, and T. J. Cox, "L[u-umlaut]ke and power residue sequence diffusers," J. Acoust. Soc. Am. **123**, 2035–2042 (2008).
[12]P. D'Antonio, "Acoustical optimization of shapes and materials used in modern architecture," J. Acoust. Soc. Am. **124**, 2563–2563 (2008).
[13]AES-4id-2001 (R2007), "AES information document for room acoustics and sound reinforcement systems—Characterization and measurement of surface scattering uniformity" (Audio Engineering Society, New York).

Why are natural sounds detected faster than pips?

Clara Suied
Department of Physiology, Development and Neuroscience, Centre for the Neural Basis of Hearing, Downing Street, Cambridge CB2 3EG, United Kingdom
clarasuied@gmail.com

Patrick Susini
Institut de Recherche et de Coordination Acoustique/Musique and Unité Mixte de Recherche 9912, Centre National de la Recherche Scientifique, 1 place Igor Stravinsky, 75004 Paris, France
patrick.susini@ircam.fr

Stephen McAdams
Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec H3A 1E3, Canada
smc@music.mcgill.ca

Roy D. Patterson
Department of Physiology, Development and Neuroscience, Centre for the Neural Basis of Hearing, Downing Street, Cambridge CB2 3EG, United Kingdom
rdp1@cam.ac.uk

Abstract: Simple reaction times (RTs) were used to measure differences in processing time between natural animal sounds and artificial sounds. When the artificial stimuli were sequences of short tone pulses, the animal sounds were detected faster than the artificial sounds. The animal sounds were then compared with acoustically modified versions (white noise modulated by the temporal envelope of the animal sounds). No differences in RTs were observed between the animal sounds and their modified counterparts. These results show that the fast detection observed for natural sounds, in the present task, could be explained by their acoustic properties.

PACS numbers: 43.66.Lj [QJF]
Date Received: November 17, 2009 **Date Accepted:** January 6, 2010

1. Introduction

The purpose of an auditory warning is to alert the user of a given system (car, plane, and hospital equipment) to a potentially dangerous situation and/or to the arrival of information on visual displays (Patterson, 1982). Several acoustical parameters have been shown to be good candidates to modulate the perceived urgency of an auditory warning: e.g., the higher the pitch and the faster the speed (in case of a multiple-burst sound), the higher the perceived urgency (Edworthy *et al.*, 1991). By contrast with these artificial auditory warnings, some authors have proposed the use of everyday sounds as warnings. For example, Graham (1999) observed shorter response times for everyday sounds (car horn and tire skid) than for conventional warnings (tone) and argued that everyday sounds are understood more quickly and easily than abstract sounds. However, simple acoustic differences, rather than semantic or cognitive differences, might be sufficient to explain the reaction-time advantage for everyday sounds.

More than an increase in the *perceived* urgency, a warning signal is efficient when it induces faster detection and increases the probability of an appropriate reaction under urgent conditions. In a companion study (Suied *et al.*, 2008), we have shown the advantages of an objective measurement [reaction time (RT)] to assess correctly the level of urgency of a sound.

In this study, we present a pair of experiments performed to investigate whether natural sounds are detected faster than artificial sounds by human listeners. First, we show that natural sounds are detected faster than simple artificial sounds (experiment 1). Then, we demonstrate that simple acoustic considerations, rather than very early recognition of the sound, can explain this behavioral advantage (experiment 2).

2. Experiment 1: Artificial sounds versus animal sounds

2.1 Methods

Twelve volunteers (7 women and 5 men; mean age 36 ± 10 years) participated in this experiment. All were naïve with respect to its purpose. None of them reported having hearing problems. The study was carried out in accordance with the Declaration of Helsinki. All participants provided informed consent to participate in the study.

Two categories of sounds were compared: classical warning sounds and animal sounds. Four sounds were tested in each category. For the classical warning sounds, we used the same template for the stimuli as in our companion paper (Suied *et al.*, 2008). The template for the different stimuli was an isochronous sequence of short pulses. Each pulse of the burst was a 1-kHz pure tone, 20 ms in duration, and included 5-ms linear onset and offset ramps. Stimuli varied along a single dimension, the interonset interval (IOI), defined as the time elapsed between the onsets of two pulses. The four IOIs tested were 100, 50, 33, and 25 ms (these four sounds are referred to hereafter as IOI100, IOI50, IOI33, and IOI25, respectively). The total duration of each burst was 220 ms. The natural sounds were animal sounds obtained from the SoundIdeas database (Sound Ideas General Series 6000, www.sound-ideas.com): a lion sound, two different leopard sounds, and one jaguar sound, referred to hereafter as Lion, Leo1, Leo2, and Jag, respectively. They were modified to be 220 ms in duration, with a linear ramp of 10 ms at the end of the sound (see Fig. 2 for the waveforms of the animal sounds).

Loudness equalization was performed on the eight stimuli to avoid any RT differences due to loudness differences (see Chocholle, 1940; Suied *et al.*, 2008). A group of nine other listeners participated in this preliminary experiment. Loudness matches were obtained with an adjustment procedure. The listener was asked to adjust the comparison stimulus until it seemed equal in loudness to the standard stimulus. IOI100 was used as the standard stimulus. The level of the standard stimulus was fixed at 76 dB sound pressure level (SPL). The mean level differences at which the comparison and the standard stimuli were judged to be equal in loudness were between 0.5 and 6 dB. IOI50 was presented at 75.5 dB SPL, IOI33 at 75.5 dB SPL, IOI25 at 75.2 dB SPL, Lion at 73.6 dB SPL, Leo1 at 75 dB SPL, Leo2 at 73.7 dB SPL, and Jag at 70 dB SPL.

The sound samples were presented at a 44.1-kHz sampling rate. They were amplified by a Yamaha P2075 stereo amplifier and presented binaurally over Sennheiser HD 250 linear II headphones. The experimental sessions were run using a Max/MSP interface on an Apple computer. Participants responded by using the space bar of the computer keyboard placed on a table in front of them. The responses were recorded by Max/MSP with a temporal precision for stimulus presentation and response collection of around 1 ms. The experiments took place in a double-walled Industrial Acoustics Co. (IAC) sound booth.

One exemplar of each of the eight stimuli was presented in random order for each trial. Following a standard simple-RT procedure, participants had to respond as soon as they detected the sound by pressing the space bar as quickly as possible. They were asked to keep the finger of their dominant hand in contact with the space bar between trials. The inter-trial interval was randomly fixed between 1 and 7 s. These stimuli were presented in six separate blocks of trials. Each block consisted of 96 stimuli. The stimuli of different IOIs were randomly intermixed. The number of stimuli of different IOIs was equal in each block (12 each), thus leading to 72 repetitions for each stimulus and each participant. Participants performed practice trials until they were comfortable with the task.

Responses were first analyzed to remove error trials, i.e., anticipations (RTs less than 100 ms) and RTs greater than 1000 ms. Each RT value was transformed to its natural logarithm

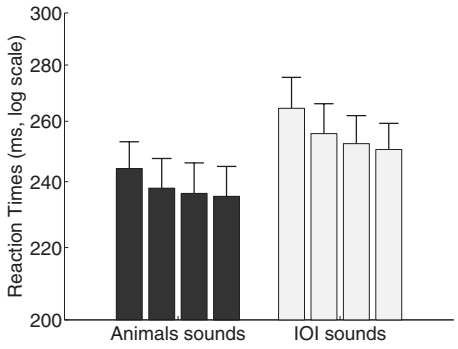


Fig. 1. RTs of the animal sounds and IOI sounds are presented (from left to right: Lion, Leo1, Leo2, Jag, IOI100, IOI50, IOI33, and IOI25; see text for details). RTs were first transformed to a log scale and then averaged across all participants. The log scale was converted back to linear ms for display purposes. The error bars represent one standard error of the mean. RTs to the animal sounds were shorter than those to the IOI sounds.

(see Ulrich and Miller, 1993; Luce, 1986), before averaging ln(RT) for each condition (see Suied *et al.*, 2009 for similar analyses on RTs). To identify between-condition differences in mean ln(RT), a repeated-measures analysis of variance (ANOVA) was conducted with sound as a within-subject factor (IOI100, IOI50, IOI33, IOI25, Lion, Leo1, Leo2, and Jag). A Kolmogorov–Smirnov test was performed to check for the normality of the distribution of residuals of the ANOVA. For this analysis, we pooled together the results for all conditions in order to increase the power of the statistical test. In addition, to account for violations of the sphericity assumption, p-values were adjusted using the Huynh–Feldt correction, and $p<0.05$ was considered to be statistically significant. Finally, we performed orthogonal contrasts to explain the main effect of the ANOVA. For the computation of the contrast in the case of repeated-measures ANOVA, the error term is based on the data on which the contrast is performed, instead of using the global error term of the ANOVA factor.

2.2 Results

There were no anticipations, only 0.2% misses and 0.2% of RTs greater than 1000 ms. These outlier data were discarded. The Kolmogorov–Smirnov test revealed that the distribution of the residuals of the ANOVA was not different from a normal distribution ($d=0.07$; $N=96$; $p>0.1$). This result validates the log transformation and shows that the original distribution of RTs was indeed log-normal.

The repeated-measures ANOVA of ln(RT) revealed a significant main effect of sound [$F(7,77)=27.25$; $\varepsilon=0.5$; $p<0.0001$]. These data are represented in Fig. 1. We then performed four mutually orthogonal contrasts [$F(4,44)=30.09$; $p<0.00001$] that show that: (1) RT was significantly shorter for the animal sounds than for the IOI sounds [Lion, Leo1, Leo2, and Jag compared to IOI100, IOI50, IOI33, and IOI25, $t(11)=6.7$; $p<0.00001$]; (2) RT was significantly longer for the Lion sound than for the three other animal sounds [$t(11)=3.5$; $p<0.005$]; (3) RT to the IOI100 sound was significantly longer than for the three other IOIs sounds [$t(11)=4.6$; $p<0.005$]; (4) RT tended to be shorter for IOI33 and IOI25 than for IOI50 [marginal significance: $t(11)=1.8$; $p=0.09$].

2.3 Discussion

Animal sounds led to a shorter RT than artificial sounds. This could be due to a very early recognition of the animal sounds. We could also hypothesize that because of some fundamental acoustical characteristic, these animal sounds induced a brainstem reflex by signaling an important and urgent event (for a review, see Juslin and Västfjäll, 2008), and this might be responsible for the shorter RT. It could also simply reflect acoustical differences, for example, in spectral

content, between the two categories of sounds: By statistical facilitation only, the greater the number of frequency channels activated, the shorter the detection process. Experiment 2 was designed to distinguish between these two possibilities.

For the IOI sounds, the shortest RTs were to IOI33. These data are consistent, at least qualitatively, with a multiple-look model for temporal integration (Viemeister and Wakefield, 1991). The IOI50 sound contains more pulses than the IOI100 sound (and similarly for the IOI33 and IOI50 sounds), so it may lead to more "looks," which might, in turn, induce shorter RTs. The threshold at 33 ms could, however, reflect another process: The lower limit of melodic pitch is around 30 Hz (Pressnitzer *et al.*, 2001). Interestingly, Russo and Jones (2007) recently found that the urgency of pulse trains is closely related to the perception of pitch: The pulse repetition rate corresponding to the transition between a pitch percept and independent pulses was judged as the most urgent and led to very short RT. For the animal sounds, the longest RT was observed for the Lion sound. This "Lion effect" will be discussed together with the results from experiment 2 below (see 3.3).

3. Experiment 2: Animal sounds versus modulated noises

In this experiment, we compared animal sounds to modified versions of the same sounds (white noise modulated with the temporal envelope of the animal sounds) in order to control for differences in spectral and temporal complexities between natural and artificial sounds in experiment 1.

3.1 Methods

Twelve new volunteers (5 women and 7 men; mean age 31 ± 7 years) participated in this experiment. All were naïve with respect to its purpose. None of them reported having hearing problems. The study was carried out in accordance with the Declaration of Helsinki. All participants provided informed consent to participate in the study.

The four animal sounds used previously in experiment 1 were tested again in experiment 2. The temporal envelopes of these sounds were applied to white noise to provide the "modulated noise" versions, denoted hereafter by the prefix "MN_." The temporal envelope was extracted using a half-wave rectifier followed by a low-pass filter (sixth-order Butterworth filter with a cut-off frequency of 5 kHz). As in experiment 1, the eight stimuli were equalized in loudness. The MN_Lion sound (used as the reference sound) was presented at 76 dB SPL, Lion at 78 dB SPL, Leo1 at 77.9 dB SPL, Leo2 at 78 dB SPL, Jag at 74.1 dB SPL, MN_Leo1 at 76 dB SPL, MN_Leo2 at 76.2 dB SPL, and MN_Jag at 75.5 dB SPL.

In addition, at the end of this second experiment, we verified that the participants could categorize the original animal sounds and their modulated noise versions correctly into "animal" and "non-animal" categories. They all performed this task very easily.

The apparatus, procedure, and statistical analyses were the same as in experiment 1.

3.2 Results

There were no anticipations, only 0.3% misses and 0.3% of RTs greater than 1000 ms. These outlier data were discarded. A Kolmogorov–Smirnov test revealed that the distribution of the residuals of the ANOVA was not different from a normal distribution ($d=0.11$; $N=96$; $p>0.1$). This result validates the log-transformation and shows that the original distribution of RTs was indeed log-normal.

The repeated-measures ANOVA on ln(RT) revealed a significant main effect of sound [$F(7,77)=6.72$; $\varepsilon=1$; $p<0.0001$]. These data are represented in Fig. 2. Three mutually orthogonal contrasts [$F(3,33)=11.62$; $p<0.00001$] showed the following: (1) There was no clear difference between RTs for the animal sounds compared to those for the MN versions [Lion, Leo1, Leo2, and Jag compared to MN_Lion, MN_Leo1, MN_Leo2, and MN_Jag, $t(11)=2.1$; $p=0.06$], and the MN sounds tended to be detected faster than the natural sounds (see Fig. 2); (2) as in experiment 1, RTs

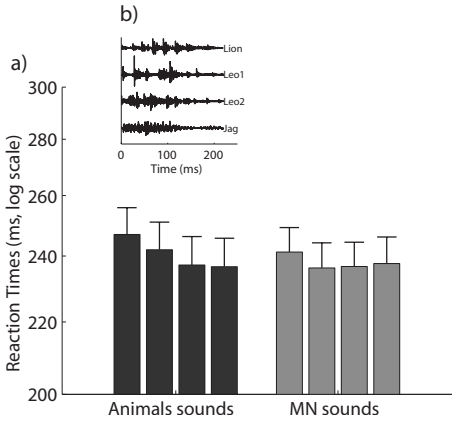


Fig. 2. (a) RTs of the animal sounds and MN sounds are presented (from left to right: Lion, Leo1, Leo2, Jag, MN_Lion, MN_Leo1, MN_Leo2, and MN_Jag; see Fig. 1 for details). RTs to the animal sounds were similar to RTs for the MN sounds that preserved the temporal envelope of the sound. (b) Temporal waveforms of the four animal sounds.

were significantly longer for the Lion sound than for the three other animal sounds [$t(11)=5.5$; $p<0.0002$]; (3) RTs were significantly longer for the MN_Lion sound than for the three other MN sounds [$t(11)=2.9$; $p<0.02$].

3.3 Discussion

We observed similar RTs for real animal sounds and their MN versions. This result validates the "acoustic" hypothesis, suggesting that the RT difference between the animal and the artificial IOI sounds in experiment 1 was indeed due to their difference in acoustic properties. Temporal and spectral differences can be responsible for the RT difference observed between the IOI sounds and the animal sounds (experiment 1). In experiment 2, similar RTs were obtained for sounds with the same temporal envelope; this suggests that differences in the temporal envelope between animal and IOI sounds could explain the faster RTs to animal sounds in experiment 1. The large difference in spectral content between repeated pure tones (IOI sounds) and animal sounds could also be responsible for the faster RTs to animal sounds. In experiment 2, we compared two categories of sounds with less obvious differences in the spectral content. If anything, there was a trend for faster RT for the MN sounds, which could be due to the higher number of channels activated for the MN sounds than for the animal sounds.

The possibility that shorter RTs for animal sounds (experiment 1) were due to cognitive factors (learned associations between feline sounds and danger, for example) is ruled out by experiment 2: RTs for animal sounds were not shorter than for the artificial MN sounds, although participants were still able to recognize animals vs non-animals sounds. Although we do not deny a plausible and potential specificity in the encoding and recognition of natural sounds, these findings suggest that, at least for simple detection tasks, the behavioral advantage for natural sounds can be easily explained by simple acoustic differences. The relationships between the acoustic characteristics of different types of animals (predators or non-predators) and RTs might be an interesting generalization of the current study.

The Lion effect observed in experiment 1 (that is, a longer RT for the Lion sound compared to the other animal sounds) was reproduced in experiment 2. Interestingly, this Lion effect held for the MN sounds, which preserved only the temporal envelope of the sounds. We computed the attack time (defined as the time it took for the temporal envelope to reach the maximum from 40 dB down) on the animal sounds; there was no obvious relationship between the attack times and the RTs that could explain the Lion effect (attack times for Lion: 96.1 ms, Leo1: 107.2 ms, Leo2: 67.7 ms, and Jag: 57.4 ms). The waveforms of the animal sounds are presented in Fig. 2. The importance of the temporal envelope for speech recognition has already

been evidenced (Shannon *et al.*, 1995). From the current data, it also seems that the temporal envelope has an impact on the speed of detection. This requires further investigation.

Acknowledgments

We would like to thank Marie Magnin and Sabine Langlois for their help. This work was partly supported by Renault SA.

References and links

Chocholle, R. (**1940**). "Variation des temps de réaction auditif en fonction de l'intensité à diverses fréquences (Variation in auditory reaction time as a function of intensity at various frequencies)," Annee Psychol. **41**, 65–124.

Edworthy, J., Loxley, S., and Dennis, I. (**1991**). "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," Hum. Factors **33**, 205–231.

Graham, R. (**1999**). "Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application," Ergonomics **42**, 1233–1248.

Juslin, P. N., and Västfjäll, D. (**2008**). "Emotional responses to music: The need to consider underlying mechanisms," Behav. Brain Sci. **31**, 559–575.

Luce, R. D. (**1986**). *Response Times: Their Role in Inferring Elementary Mental Organization* (Oxford University Press, New York).

Patterson, R. D. (**1982**). "Guidelines for auditory warning systems on civil aircraft," Civil Aviation Authority Paper No. 82017.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Russo, F. A., and Jones, J. A. (**2007**). "Urgency is a non-monotonic function of pulse rate," J. Acoust. Soc. Am. **122**, EL185–EL190.

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

Suied, C., Bonneel, N., and Viaud-Delmon, I. (**2009**). "Integration of auditory and visual information in the recognition of realistic objects," Exp. Brain Res. **194**, 91–102.

Suied, C., Susini, P., and McAdams, S. (**2008**). "Evaluating warning sound urgency with reaction times," J. Exp. Psychol., Appl. **14**, 201–212.

Ulrich, R., and Miller, J. (**1993**). "Information processing models generating lognormally distributed reaction times," J. Math. Psychol. **37**, 513–525.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.

The behavioral audiogram of whitetail deer (*Odocoileus virginianus*)

Henry Heffner, Jr. and Henry E. Heffner[a)]
Whitetail Deer Research LLC, P.O. Box 412, Swanton, Ohio 43558
hheff@adelphia.net, heff270@yahoo.com

Abstract: The behavioral audiograms of two female white-tailed deer (*Odocoileus virginianus*) were determined using a conditioned-suppression avoidance procedure. At a level of 60 dB sound pressure level, their hearing range extends from 115 Hz to 54 kHz with a best sensitivity of −3 dB at 8 kHz; increasing the intensity of the sound extends their hearing range from 32 Hz (at 96.5 dB) to 64 kHz (at 93 dB). Compared with humans, white-tailed deer have better high-frequency but poorer low-frequency hearing.

PACS numbers: 43.80.Lb, 43.66.Gf, 43.66.Cb [CM]
Date Received: November 10, 2009 **Date Accepted:** December 10, 2009

1. Introduction

White-tailed deer have perhaps the largest economic impact of any wild animal in North America. On the one hand, billions of dollars are spent each year on equipment and travel related to deer hunting (e.g., Conover, 1997). On the other hand, deer spread Lyme disease, causing millions of dollars of damage to both the agriculture and timber industries, and over a billion dollars of damage to vehicles that collide with them each year (Conover, 1997; Schwabe and Schuhmann, 2002). Indeed, more than 100 people are killed each year in deer-vehicle collisions, with many more seriously injured, making deer the most deadly wild animals in North America (Bailey, 2001). As a result, there is much interest in the behavior of white-tailed deer, particularly in their sensory abilities (e.g., Gerlach *et al.*, 1994).

Because deer are naturally afraid of humans and do not readily tolerate our presence, it is difficult to conduct behavioral tests on them. As a result, the only measure of their hearing currently available is their auditory brainstem response (D'Angelo *et al.*, 2007). This measurement indicates that the hearing range of deer extends from 250 Hz to 30 kHz, with a best sensitivity of only 42 dB at 4 and 8 kHz. However, the auditory brainstem response does not give an accurate measure of an animal's absolute sensitivity nor does it necessarily indicate the relative sensitivity of an animal to different frequencies (Heffner and Heffner, 2003). Such information can only be obtained from a behavioral audiogram.

We present here the absolute pure-tone thresholds obtained for two domestically raised whitetail deer using standard animal psychophysical procedures. These results will be of interest to those who wish to attract, repel, or conceal their presence from deer.

2. Method

The animals used in this experiment were two whitetail does (*Odocoileus virginianus*) 1–2 years of age that had been born and raised domestically. The animals were weighed daily during testing to help monitor their health.

The deer were tested using a conditioned-suppression avoidance procedure in which a thirsty animal was trained to maintain mouth contact with a small stainless steel water bowl in order to receive a steady trickle of water. The bowl, which was mounted on a post 0.5 m above the floor, was carefully positioned so that the animals made little or no drinking noise. Pure tones were then presented at random intervals followed by a mild electric shock delivered between the

[a)]Author to whom correspondence should be addressed.

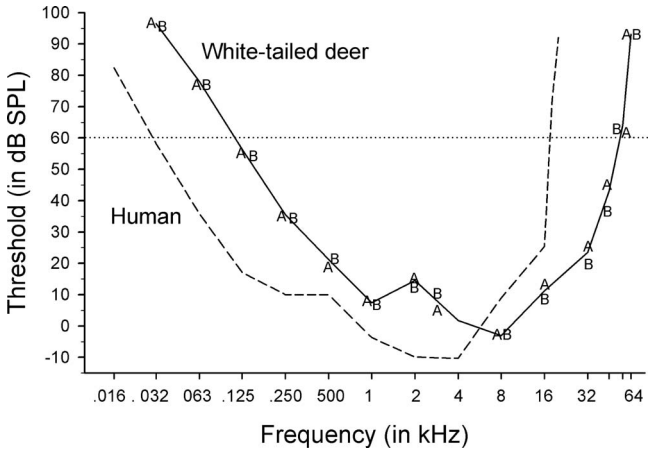


Fig. 1. Absolute thresholds of two white-tailed deer (A and B). The solid line indicates the average thresholds of the deer. The dashed line is a human audiogram obtained in comparable free-field conditions (Jackson *et al.*, 1999). The horizontal dotted line indicates the 60 dB sound pressure level, which is the level commonly used when comparing the hearing ranges of different species.

bowl and the floor. An animal avoided the shock by breaking contact with the bowl whenever it heard a tone, thereby also indicating that it had detected the tone. Thus, the task resembles the natural situation in which an animal at a water hole pauses when it senses danger. That the shock level used was "mild" was indicated by the fact that the deer never developed a fear of the water bowl and returned to it as soon as the sound was turned off (for details of this procedure, see Heffner and Heffner, 1995).

Thresholds were determined by reducing the intensity of a tone in 5 dB steps until an animal could no longer detect it above chance level. Threshold was defined as the intensity at which an animal could detect a sound 50% of the time (corrected for false positives), which was usually calculated by interpolation (Heffner and Heffner, 1995).

To produce the tones, sine waves were generated with an oscillator or a digital signal generator, gated with a rise-fall gate set to at least 10 ms to eliminate onset and offset artifacts, pulsed (400 ms on, 100 ms off), filtered ($\pm 1/3$ octave settings), monitored with an oscilloscope, amplified, and sent to a loudspeaker: either a subwoofer (32–63 Hz), woofer (125–4 kHz), piezoelectric tweeter (8–32 kHz), or leaf tweeter (45, 56, and 64 kHz). Most testing was conducted with the speaker placed in front of the animal at a distance of 1 m. However, because the animals sometimes did not point their pinnae forward, a second speaker was placed off to one side or behind an animal at a distance of 1.5 m or more with the exact location depending on the direction in which an animal tended to point its pinna. The procedure for calibrating the sound has been described elsewhere (Jackson *et al.*, 1997) and the sound pressure level (SPL) used was referenced to 20 $\mu N/m^2$. Testing was conducted in a single-wall sound-proof chamber, the walls and ceiling of which were lined with acoustic foam.

3. Results

The deer learned to enter the sound chamber, drink from the water bowl, and break contact with the bowl whenever a suprathreshold stimulus was presented. Although the animals usually pointed their pinnae straight ahead at the loudspeaker located in front of them, they sometimes rotated their pinnae to the side or back as though they were checking for sounds coming from those directions. When this occurred, the tone trials were delayed until their pinnae were again directed toward the loudspeaker. However, for frequencies of 8 kHz and higher, a second loudspeaker was placed in the location toward which they oriented their pinna. This procedure resulted in stable thresholds that represent the animals' optimal sensitivity.

The audiograms of the two deer, shown in Fig. 1, have the characteristic shape of

mammalian audiograms. Beginning at the low frequencies, the deer were able to hear 32 Hz at an average threshold of 96.5 dB with sensitivity improving as frequency was increased. The animals showed a broad range of good sensitivity extending from 500 Hz to 32 kHz with a best threshold of −3 dB at 8 kHz. Above 32 kHz, sensitivity decreased rapidly to an average threshold of 93 dB at 64 kHz. Overall, at an intensity of 60 dB SPL, the deer were able to hear from 115 Hz to 54 kHz.

4. Discussion

There are two points to note about this deer audiogram. First, because sounds were only presented when at least one, if not both, of an animal's pinnae were pointed toward a loudspeaker, they represent the animals' optimal sensitivity. As has been demonstrated in reindeer, the pinnae of deer are directional for high frequencies and sensitivity may be reduced by 20 dB or more when the pinnae are pointed away from the sound source (Flydal *et al.*, 2001). Second, although we only tested females, these results are expected to apply to male deer as well, as the auditory sensitivity of mammals has not been observed to differ between the sexes (Heffner and Heffner, 2003).

Because of the extensive interactions between humans and white-tailed deer, it is of interest to compare their audiogram with that of humans. As can be seen in Fig. 1, the deer audiogram is similar in shape to that of humans and, indeed, looks like the human audiogram shifted approximately two octaves toward the higher frequencies. Some of the differences between the human and deer audiogram are well understood. In particular, the better high-frequency hearing of deer is explained by the observation that mammals rely on high-frequency cues to localize sound, high frequencies being particularly important for localization in the vertical plane and for preventing front-back confusions (Heffner and Heffner, 2008). However, because the directionality of high frequencies depends on the size of an animal's head and pinnae, the smaller the animal, the higher it must hear in order to use the high-frequency locus cues. Thus, deer hear higher than humans because they are smaller. However, less is known about the variation in mammalian low-frequency hearing and there is currently no explanation for the difference in the low-frequency sensitivity of humans and deer (Heffner and Heffner, 2003). (The audiograms of other mammals are available at http://psychology.utoledo.edu/showpage.asp?name=mammal_hearing for comparison with deer.)

Finally, the audiogram provided here can be used to obtain a preliminary estimate of the audibility of a sound to deer. That is, sounds whose spectra fall within the bounds of the audiogram will be audible to deer if they reach a deer's ear at a sufficient sound pressure level. Sounds that fall above or below the frequency range of deer will not be audible to them regardless of the intensity. However, care must be taken in using these data to estimate the relative loudness of sound to deer. This is because the audiogram measures the sensitivity to pure tones, whereas most sounds of interest are complex sounds containing multiple frequencies, and it has been shown that the perceived loudness of such sounds is not easily estimated from measures of pure-tone sensitivity (e.g., Hellman and Zwicker, 1987).

References and links

Bailey, R. (**2001**). "North America's most dangerous mammal," Reason, http://reason.com/archives/2001/11/21/north-americas-most-dangerous/print (Last viewed 11/27/2009).
Conover, M. R. (**1997**). "Monetary and intangible valuation of deer in the United States," Wildlife Soc. B **25**, 298–305.
D'Angelo, G. J., De Chicchis, A. R., Osborn, D. A., Gallagher, G. R., Warren, R. J., and Miller, K. V. (**2007**). "Hearing range of white-tailed deer as determined by auditory brainstem response," J. Wildl. Manage. **71**, 1238–1242.
Gerlach, D., Atwater, S., and Schnell, J. (**1994**). *Deer* (Stackpole Books, Mechanicsburg, PA).
Flydal, K., Hermansen, A., Enger, P. S., and Reimers, E. (**2001**). "Hearing in reindeer (*Rangifer tarandus*)," J. Comp. Physiol., A **187**, 265–269.
Heffner, H. E., and Heffner, R. S. (**1995**). "Conditioned avoidance," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser, Basel), pp. 73–87.
Heffner, H. E., and Heffner, R. S. (**2003**). "Audition," in *Handbook of Research Methods in Experimental Psychology*, edited by S. F. Davis (Blackwell, Malden, MA), pp. 413–440.
Heffner, H. E., and Heffner, R. S. (**2008**). "High-frequency hearing," in *Handbook of the Senses: Audition*, edited by P. Dallos, D. Oertel, and R. Hoy (Elsevier, New York), pp. 55–60.
Hellman, R., and Zwicker, E. (**1987**). "Why can a decrease in dB(A) produce an increase in loudness?," J.

Acoust. Soc. Am. **82**, 1700–1705.
Jackson, L. L., Heffner, H. E., and Heffner, R. S. (**1997**). "Audiogram of the fox squirrel (*Sciurus niger*)," J. Comp. Psychol. **111**, 100–104.
Jackson, L. L., Heffner, R. S., and Heffner, H. E. (**1999**). "Free-field audiogram of the Japanese macaque (*Macaca fuscata*)," J. Acoust. Soc. Am. **106**, 3017–3023.
Schwabe, K. A., and Schuhmann, P. W., (**2002**). "Deer-vehicle collisions and deer value: An analysis of competing literatures," Wildlife Soc. B **30**, 609–615.

Dispersion of sound in dilute suspensions with nonlinear particle relaxation

Max Kandula
ASRC Aerospace, NASA Kennedy Space Center, Florida 32899
max.kandula-1@ksc.nasa.gov

Abstract: The theory accounting for nonlinear particle relaxation (viscous and thermal) has been applied to the prediction of dispersion of sound in dilute suspensions. The results suggest that significant deviations exist for sound dispersion between the linear and nonlinear theories at large values of $\omega\tau_d$, where ω is the circular frequency and τ_d is the Stokesian particle relaxation time. It is revealed that the nonlinear effect on the dispersion coefficient due to viscous contribution is larger relative to that of thermal conduction.

PACS numbers: 43.20.Hq, 43.50.Nm, 43.50.Gf [AN]
Date Received: October 12, 2009 **Date Accepted:** January 6, 2010

1. Introduction

Sound attenuation in fluids, representing the dissipation of acoustic energy from a sound wave, occurs through a number of physical processes involving molecular viscosity, thermal conductivity, and other dissipative or relaxation processes.[1–7] When a fluid contains inhomogeneities such as suspended particles (solid particles, drops, and bubbles), additional viscous and heat conduction losses occur in the immediate neighborhood of the suspended particles.[1–3,8] Particle relaxation in a dilute suspension is the process by which particles adjust to fluctuations (velocity and temperature) of the surrounding fluid, leading to attenuation and dispersion of a sound wave. A comprehensive review of the physics and scientific history of acoustic interactions with particulate mixtures is provided by Challis *et al.*[9] The acoustic intensity I of a plane wave propagating through an absorbing medium is expressed by

$$I = I_0 e^{-\alpha_i x}, \tag{1}$$

where x is the distance traversed, I_0 is the intensity at $x=0$, and α_i is the intensity attenuation coefficient for the medium. The quantity α_i depends on viscosity, thermal conductivity, and other factors such as molecular relaxation.

Sound propagation in aerosols and fog has been studied experimentally and theoretically by several investigators since the pioneering work of Sewell,[10] with the aid of a scattering formulation on the assumption of immovable particles. Epstein[11] extended this theory for particles in motion, and Epstein and Carhart[12] additionally considered heat conduction effects. Allegra and Hawley[13] provided further extensions by including liquid-liquid as well as liquid-solid systems.

Temkin and Dobbins,[14] in their classical work involving a coupled-phase formulation, theoretically considered particle attenuation and dispersion of sound in a manner which illustrates explicitly the relaxation character of the problem. Basset history and added mass terms were included in an elegant coupled-phase formulation by Harker and Temple.[15] Coupled phase effects were also treated by Evans and Attenborough[16,17] in an extension to the work of Harker and Temple[15] to incorporate thermal conduction.

The absorption of sound in suspensions of irregular (nonspherical) particles was considered by Urick.[18] Experimental and theoretical studies (extension of Urick's model) on sound absorption involving irregular particles were also considered by Richards *et al.*[19,20]

The particulate relaxation models for the sound attenuation are all based on Stokes drag (linear drag law) and pure conduction limit (linear heat transfer). Recently the author[21] investigated sound attenuation in dilute suspensions and extended the theory of Temkin and

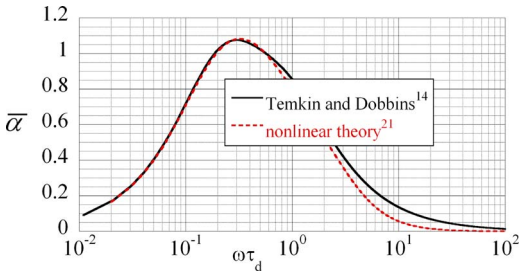


Fig. 1. (Color online) Predicted absorption coefficient with nonlinear particle relaxation processes.

Dobbins[14] by considering nonlinear drag and heat transfer laws applicable to relatively large-sized droplets. The absorption coefficient per unit frequency predicted by the nonlinear theory is compared with that indicated by the theory of Temkin and Dobbins[14] in Fig. 1. In Fig. 1,

$$\bar{\alpha} = c_0\alpha/(C_m\omega) = f(\omega\tau_d, \mathrm{Pr}, c_{pp}/c_{pg}, \gamma), \tag{2a}$$

where

$$\tau_d = d_p^2\rho_p/(18\mu_g) \tag{2b}$$

and

$$C_m = n_0 m_p/\rho_g. \tag{2c}$$

In the above, $\bar{\alpha}$ is the attenuation per unit frequency per unit mass fraction, α is the amplitude attenuation coefficient, c_0 is the speed of sound in the gas phase, ω is the circular frequency, C_m is the mass concentration, τ_d is the dynamic relaxation time of the particle (relating to particle-fluid velocity lag), n_0 is the mean number of particles per unit volume of mixture, m_p is the mass of one particle, ρ_p is the mean particle density, ρ_g is the mean density of gas, μ_g is the mean dynamic viscosity of gas, and d_p is the particle diameter. Also the quantity Pr refers to Prandtl number of the gas, c_{pg} is the specific heat of gas, c_{pp} is the specific heat of the particle, and γ is the isentropic exponent (specific-heat ratio). Note that $\alpha = \alpha_i/2$. The results shown in Fig. 1 correspond to $c_{pp}/c_{pg} = 4.17$, $\mathrm{Pr} = 0.71$, and $\gamma = 1.4$ (representative of water droplets in air[14]). With the aid of this nonlinear model, the existence of the spectral peak in the linear absorption coefficient α has been demonstrated.[21]

Based on this extension, good agreement was achieved with the recent data of Norum[22] for sound attenuation in perfectly expanded supersonic jets containing suspended water droplets, which reveal that the linear absorption coefficient displays a spectral peak (Fig. 2, adapted from Ref. 21). The data correspond to hot supersonic jet of air from a convergent-divergent nozzle operation at a jet total temperature $T_t = 867$ K and a jet exit Mach number M_j

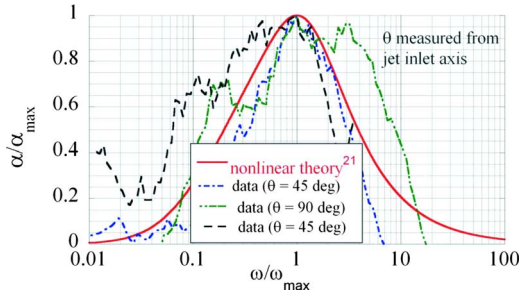


Fig. 2. (Color online) Comparison of the predictions for the linear absorption coefficient with test data of Norum (Ref. 22).

=1.45. The jet Mach number is defined as $M_j = u_j / c_j$, where the subscript j refers to the nozzle exit conditions. The mass flow rate (maximum considered) of water to that of the jet is about 0.85. The angle θ is measured from the jet inlet axis. In the data, water is injected at 45°.

Similar spectral peaks in the attenuation coefficient have been observed in the measurements by Krothapalli *et al.*[23] for microjet injection of water into high speed exhaust jets. Spectral peaks in the linear absorption coefficient in dilute suspensions were also noticed in the experimental data of Richards *et al.*[19,20] covering a frequency range 50–150 kHz.

In the preceding work by the author,[21] results were presented only for sound attenuation with nonlinear particle relaxation, and the effect of nonlinearity on the dispersion effects was not considered. In practical applications such as combustion chamber instability (related to liquid propellants or metallized solid propellants) and exhaust jets with injected water droplets, in addition to sound attenuation (damping of pressure oscillations) the dispersion effects (phase velocity changes) are also important from the standpoint of sound propagation. The present work applies the nonlinear theory for the dispersion of sound in dilute suspensions and compares the linear and nonlinear theories for sound dispersion.

2. Sound dispersion

The present analysis for sound dispersion is similar to that proposed by the author[21] for sound attenuation and will be briefly presented as below. Without any loss of generality, the dispersion of sound for large particle Reynolds numbers with nonlinear particle relaxation may be expressed with the aid of Temkin and Dobbins[14] results as follows:

$$\bar{\beta} = \left[\left(\frac{c_0}{c} \right)^2 - 1 \right] \Big/ \; C_m = \frac{1}{1 + \omega \tau_{d1}^2} + (\gamma - 1) \left(\frac{c_{pp}}{c_{pg}} \right) \frac{1}{1 + \omega^2 \tau_{t1}^2}. \tag{3}$$

In the above the quantity, $\bar{\beta}$ is the dimensionless dispersion coefficient, c_p is the specific heat, c is the actual speed of sound in two-phase medium (phase velocity), and γ is the isentropic exponent (specific-heat ratio). The subscripts g and p, respectively, denote the gas and the particle.

The relaxation times τ_{d1} and τ_{t1} correspond to those under nonlinear drag conditions (generally representative of large-sized particles). Physically the dynamic relaxation time τ_{d1} is a measure of the time scale in which the particles follow (respond to) the fluctuations in the fluid motion.[3] Likewise, the thermal relaxation time τ_{t1} is a measure of the thermal response time of the particles to follow the fluctuations in the temperature of the fluid. They are related to the relaxation times τ_d and τ_t by the relations[21]

$$\tau_{d1} = \tau_d \psi_1(\mathrm{Re}_p), \quad \tau_{t1} = \tau_t \psi_2(\mathrm{Re}_p, \mathrm{Pr}), \tag{4a}$$

where

$$\psi_1(\mathrm{Re}_p) = C_{D1}/C_D, \quad \psi_2(\mathrm{Re}_p, \mathrm{Pr}) = \mathrm{Nu}_1/\mathrm{Nu}, \tag{4b}$$

with C_{D1} standing for the nonlinear drag coefficient and Nu_1 for the nonlinear heat transfer (Nusselt number). In the above, the quantity τ_d is given by Eq. (2b), and

$$\tau_t = \frac{m_p c_{pp}}{2\pi d_p k_g} = \frac{\mathrm{Pr}\, c_{pp} d_p^2 \rho_p}{12 \mu_g c_{pg}} = \left(\frac{3}{2} \right) \left(\frac{c_{pp}}{c_{pg}} \right) \mathrm{Pr}\ \tau_d. \tag{4c}$$

Also the particle Reynolds number Re_p is defined by

$$\mathrm{Re}_p = \rho_g |u_g - u_p| d_p / \mu_g, \tag{4d}$$

where u_g and u_p denote the velocity of the gas and the particle, respectively. Also the quantity $\mathrm{Pr} = c_{pg} \mu_g / k_g$ stands for the Prandtl number of gas, where k_g stands for the thermal conductivity of the gas.

The drag coefficient and the Nusselt number in Eq. (4b) are defined by

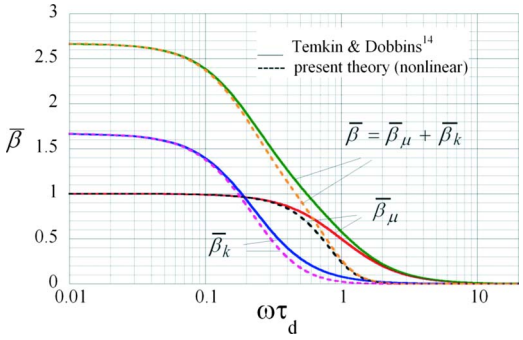


Fig. 3. (Color online) Comparison of the predictions for the dispersion coefficient between the linear and nonlinear theories for particle relaxation.

$$C_D = 2F_p \Big/ \left(\rho_g \frac{\pi}{4} d_p^2 u_g^2\right), \quad \mathrm{Nu} = h_g d_p / k_g, \tag{5}$$

where F_p is the particle drag force, and h_g is the gas-droplet convective heat transfer coefficient. For rigid particles, the linear droplet drag and heat transfer are, respectively, obtained from

$$F_p = 6\pi\mu_g(u_p - u_g), \tag{6a}$$

$$Q_p = 2\pi d_p k_g (T_p - T_g), \tag{6b}$$

which correspond to the zero droplet Reynolds number limit ($\mathrm{Re}_p \to 0$). According to Temkin and Dobbins,[14] Stokes linear drag law can be justified for $0 \le \omega\tau_d \approx 1$, provided that $\rho_g/\rho_p \ll 1$ and $(\omega d_p^2/8\nu_g)^{1/2} \ll 1$.

The expressions for ψ_1 and ψ_2 have been taken as[21]

$$\psi_1(\mathrm{Re}_p) = 1 + \frac{\mathrm{Re}_p}{24}\left(\frac{6}{1+\sqrt{\mathrm{Re}_p}} + 0.4\right), \tag{7}$$

$$\psi_2(\mathrm{Re}_p, \mathrm{Pr}) = \mathrm{Nu}_1/\mathrm{Nu} = 1 + 0.3\ \mathrm{Re}_p^{0.5}\ \mathrm{Pr}^{0.33}. \tag{8}$$

Equation (7) is obtained from Ref. 24, and Eq. (8) is based on Ref. 25.

The determination of particle Reynolds number required in the evaluation of the functions ψ_1 and ψ_2 in Eqs. (7) and (8), respectively, is exceedingly complex. There exists relatively little information on the dependence of particle Reynolds number on the particle characteristics in two-phase flows. In this connection the author[21] postulated that the particle Reynolds number depends only on the particle relaxation time and is independent of the particle to fluid density ratio for large particle to fluid density ratio, and the following power law relation is proposed based on the work of Ref. 26:

$$\mathrm{Re}_p = f(\omega\tau_d) = c(\omega\tau_d)^3. \tag{9}$$

The adjustable constant c is determined from a correlation of the theory with the test data. A value of $c = 10$ was found to be satisfactory based on the data of Norum[21] for water droplets in a supersonic air jet (Fig. 2).

3. Results and comparison

Figure 3 illustrates a comparison of the predictions for the dispersion coefficient between the linear and nonlinear theories for particle relaxation. The results are shown for $c_{pp}/c_{pg} = 4.17$, $\mathrm{Pr} = 0.71$, and $\gamma = 1.4$ (representative of water droplets in air[14]). The contributions of viscosity and the thermal conductivity along with their combined effect on the dispersion coefficient are shown for

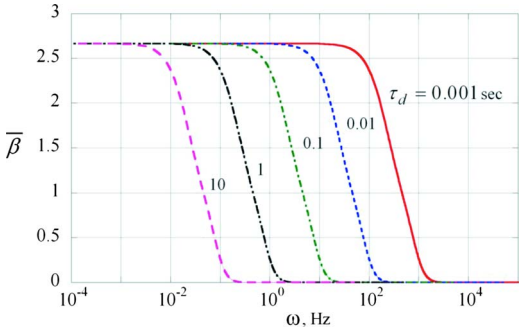


Fig. 4. (Color online) Variation in dispersion coefficient with frequency for various values of particle relaxation time.

both linear and nonlinear relaxations. It is seen that the nonlinear effects become important for $\omega\tau_d > 0.2$ (where heat conduction effects become important). With regard to the viscous contribution, nonlinearities are manifested for somewhat higher values of $\omega\tau_d$ in excess of about 0.4. The results also suggest that the nonlinear effect is more significant in the viscous contribution relative to that of thermal conduction. With nonlinear particle relaxation, the total dispersion coefficient approaches zero for smaller values of $\omega\tau_d$ than those in the case of linear relaxation.

The predictions for the dependence of the dispersion coefficient (yielding the phase velocity c) on the frequency for various values of the dynamic relaxation time τ_d is presented in Fig. 4. For example, the dynamic relaxation time τ_d for silica particles in air is about 10^{-3} s for a 5 μm particle and about 10^{-1} s for a 50 μm particle.[8] On the other hand, the thermal relaxation times for particles in air are about 5×10^{-4} s for a 5 μm particle and about 5×10^{-2} s for a 50 μm particle. For comparison purposes, the molecular and thermal relaxation time scales are of the order of 10^{-10} s for gases. The results suggest that as the particle relaxation time increases, the dispersion curve is shifted to lower frequencies, as is to be expected.

It is believed the nonlinear particle relaxation effects on sound attenuation and dispersion will be of interest in the prediction of jet noise reduction by water injection.[27–29]

4. Conclusion

The theory of nonlinear particle relaxation proposed previously for sound attenuation in dilute suspensions has been extended to investigate the effect of nonlinearity on sound dispersion. The results reveal that significant nonlinear effects are noticed at relatively large particle relaxation times. It is also observed that the nonlinear effect on dispersion due to viscous contribution is larger relative to that of thermal conduction.

Acknowledgment

Thanks are due to the reviewer for helpful suggestions in improving the manuscript. Papers by the author from 1973 to 1982 were published with the name K. Mastanaiah.

References and links

[1]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[2]A. D. Pierce, *Acoustics—An introduction to Physical Principles and Applications* (The Acoustical Society of America, New York, 1994).
[3]S. Temkin, *Elements of Acoustics* (The Acoustical Society of America, New York, 2001).
[4]J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. **2**.
[5]G. G. Stokes, "On the theories of internal friction of fluids in motion," Trans. Cambridge Philos. Soc. **8**, 287–305 (1845).
[6]H. von Helmholtz, "On the influence of friction in the air on sound motion," Verhandl. Naturhist. Med. Ver. Heidelberg **3**, 16–20 (1863); reprinted in Wissenschftliche Abhandlungen, Barth, Leipzig 1, 383–387 (1882).
[7]G. Kirchhoff, "On the influence of heat conduction in a gas on sound propagation," Ann. Phys. Chem. **134**, 177–193 (1868).

[8]S. Temkin, *Suspension Acoustics* (Cambridge University Press, Cambridge, 2005).
[9]R. E. Challis, M. J. W. Povey, M. L. Mather, and A. K. Holmes, "Ultrasound techniques for characterizing colloidal dispersions," Rep. Prog. Phys. **68**, 1541–1637 (2005).
[10]C. J. T. Sewell, "On the extinction of sound in a viscous atmosphere by small obstacles of cylindrical and spherical form," Philos. Trans. R. Soc. London, Ser. A **210**, 239–270 (1910).
[11]P. S. Epstein, "On the absorption of sound waves in suspensions and emulsions," *Theodore von Karman Anniversary Volume* (California Institute of Technology, Pasadena, CA, 1941), pp. 162–188.
[12]P. S. Epstein and R. R. Carhart, "The absorption of sound in suspensions and emulsions, I. Water fog in air," J. Acoust. Soc. Am. **25**, 553–565 (1953).
[13]J. R. Allegra and S. A. Hawley, "Attenuation of sound in suspensions and emulsions: Theory and experiments," J. Acoust. Soc. Am. **51**, 1545–1564 (1972).
[14]S. Temkin and R. A. Dobbins, "Attenuation and dispersion of sound by particulate relaxation processes," J. Acoust. Soc. Am. **40**, 317–324 (1966).
[15]A. H. Harker and J. A. G. Temple, "Velocity and attenuation of ultrasound in suspensions of particles in fluid," J. Phys. D: Appl. Phys. **21**, 1576–1588 (1988).
[16]J. M. Evans and K. Attenborough, "Coupled phase theory for sound propagation in emulsions," J. Acoust. Soc. Am. **102**, 278–282 (1997).
[17]J. M. Evans and K. Attenborough, "Sound propagation in concentrated emulsions: Comparison of coupled phase model and core-shell model," J. Acoust. Soc. Am. **112**, 1911–1917 (2002).
[18]R. J. Urick, "The absorption of sound in suspensions of irregular particles," J. Acoust. Soc. Am. **20**, 283–289 (1948).
[19]S. D. Richards, T. G. Leighton, and N. R. Brown, "Sound absorption by suspensions of nonspherical particles: Measurements compared with predictions using various particle sizing techniques," J. Acoust. Soc. Am. **114**, 1841–1850 (2003).
[20]S. D. Richards, T. G. Leighton, and N. R. Brown, "Visco-inertial absorption in dilute suspensions of irregular particles," Proc. R. Soc. London, Ser. A **459**, 2153–2167 (2003).
[21]M. Kandula, "Spectral attenuation of sound in dilute suspensions with nonlinear particle relaxation," J. Acoust. Soc. Am. **124**, EL284–EL290 (2008).
[22]T. D. Norum, "Reductions in multi-component jet noise by water injection," in 10th AIAA/CEAS Aeroacoustics Conference, Manchester, Great Britain (2004), Paper No. AIAA-2004-2976.
[23]A. Krothapalli, L. Venkatakrishnan, L. Lourenco, B. Greska, and R. Elavarasan, "Turbulence and noise suppression of a high-speed jet by water injection," J. Fluid Mech. **491**, 131–159 (2003).
[24]F. M. White, *Viscous Fluid Flow*, 2nd ed. (McGraw-Hill, New York, 1991).
[25]W. E. Ranz and W. P. Marshall, "Evaporation from drops, Part I," Chem. Eng. Prog. **48**, 141–146 (1952); "Evaporation from drops, Part II," Chem. Eng. Prog. **48**, 173–180 (1952).
[26]E. N. Ganic and K. Mastanaiah, "Investigation of droplet deposition from a turbulent gas stream," Int. J. Multiphase Flow **7**, 401–422 (1981).
[27]M. Kandula, "Prediction of turbulent jet mixing noise reduction," AIAA J. **46**, 2714–2722 (2008).
[28]M. Kandula, "Broadband shock noise reduction in turbulent jets by water injection," Appl. Acoust. **70**, 1009–1014 (2009).
[29]M. Kandula, "On the scaling laws and similarity spectra for jet noise in subsonic and supersonic flow," Int. J. Acoust. Vib. **13**, 3–16 (2008).

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.

Causal analysis of the viscoelastic Lamb problem (L)

André Moura[a)]
Institute of Physical Chemistry, University of Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany

(Received 25 April 2009; revised 26 October 2009; accepted 11 December 2009)

A mathematical development is given for the generation of viscoelastic waves by an impulsive line source acting on the interface of a viscoelastic half space, where the viscoelasticity is characterized by two relaxation processes. The considered idealized viscoelastic medium is isotropic and characterized by two Lamé constants appropriate for low frequencies, by their increments associated with the shift from low to high frequencies, and by separate relation times associated with each of the Lamé constants. A causal solution is developed using integral transforms and an extension of Cagniard's method [DOI: 10.1121/1.3290981]

PACS number(s): 43.35.Mr, 43.20.Bi, 43.20.Px [WMC] Pages: 1185–1188

I. INTRODUCTION

The viscoelastic Lamb problem is solved by means of a natural extension of the method we used to solve one-wave propagation problem,[1] which is based on the works of Ludvig and Levin[2] and the approach initiated by Strick.[3]

Lamb originally solved the homogeneous isotropic elastic half space problem for an impulse normal load acting on the free surface in order to analyze the Rayleigh wave contribution to the displacement field, i.e., the famous well-known surface wave in seismology theoretically predicted by Rayleigh in the framework of elastodynamics. Lamb's problem now refers to general source/medium geometries with a single interface and the main interest is the exact calculation in the time domain of waves emanating from an impulsive line or point source.

For ultrasonic or seismic waves, "friction" refers to the overall sum of all mechanisms responsible for the lost of elastic potential energy. We restrain our self to the losses due to viscosity because of the existence of a well-known mathematical procedure that conserves the causal aspect from the elastic to the viscoelastic problem, i.e., the so-called viscoelastic correspondence principle. Thus, consider the medium of propagation as a viscoelastic homogeneous continuum, and the problem is formulated in the framework of linear continuum mechanics in Eulerian representation. For the sake of simplicity and without lost of generality, we consider a two-dimensional problem instead of a three-dimensional one, i.e., the homogeneous isotropic viscoelastic half space problem for an infinite impulse normal line load acting on the free surface.

[a)]Electronic mail: moura@uni-heidelberg.de

II. PROBLEM FORMULATION

Let us consider a viscoelastic homogeneous half space V to the sense of the continuum mechanics. The outer of the half space is supposed to be occupied by vacuum. Hence waves can only propagate in the half space. In order to write the boundary conditions, the points of the space are represented in a Cartesian coordinate system $(Oxyz)$ with the Oxy-plane defining the free surface S and the Oz-axis oriented toward the interior of V. Application of an infinite spatially uniform line load over the free surface renders the problem two dimensional, and for the sake of simplicity and without lost of generality, one will orient it along the Oy-axis. Moreover, because of the problem symmetry with respect to the Oz-axis, only positive values of x are considered.

The elastic properties are defined by the Lamé constants λ_0 and μ_0, and the related relaxation functions for the special case of linear viscoelastic solid write, respectively,[4]

$$\lambda(t) = \frac{\lambda_0}{\lambda_0 + \lambda_1}(\lambda_1 + \lambda_0 \exp(-t/\tau_0)) \quad \text{and}$$

$$\mu(t) = \frac{\mu_0}{\mu_0 + \mu_1}(\mu_1 + \mu_0 \exp(-t/\omega_0)), \tag{1}$$

where λ_o (μ_o), λ_1 (μ_1), and τ_o (ω_o) are constants, with the first corresponding to the Lamé elastic constant and the last to the relaxation time with t, the time. The dilatational and shear elastic wave velocities then become

$$c_d = \sqrt{(\lambda_0 + 2\mu_0)/\rho} \quad \text{and} \quad c_s = \sqrt{\mu_0/\rho}, \tag{2}$$

respectively, with ρ, the mass density, from which the velocities are deduced from the two following wave equations in the elastic case:

$$\partial_{xx}^2\varphi+\partial_{zz}^2\varphi=\frac{1}{c_d^2}\partial_{tt}^2\varphi \quad \text{and} \quad \partial_{xx}^2\psi+\partial_{zz}^2\psi=\frac{1}{c_s^2}\partial_{tt}^2\psi, \tag{3}$$

where φ and ψ are the scalar potential and the nonvanishing component of the vector potential, respectively, and the operator ∂ denotes the derivative. For instance, ∂_{xx}^2 means the second-order spatial derivative with respect to x.

Because we restrain our self to the sake of the Green's function, one considers an impulsive load, and writes both the source and the boundary-initial conditions on S together as follows:

$$\partial_t(\lambda(t)(\partial_{xx}^2\varphi+\partial_{zz}^2\varphi)+2\mu(t)(\partial_{zz}^2\varphi+\partial_{xz}^2\psi))=-F\delta(t)\delta(x), \tag{4}$$

$$2\partial_{xz}^2\varphi+\partial_{zz}^2\psi-\partial_{xx}^2\psi=0, \tag{5}$$

where δ denotes the Dirac delta function, the asterisk represents the time convolution, and the constant F represents a force per unit length. Additionally, we impose the following radiation conditions in order to work only with bounded physical functions:

$$\lim_{z\to+\infty}\varphi(x,z,t)=\lim_{z\to+\infty}\psi(x,z,t)=0. \tag{6}$$

III. APPLICATION OF SPECIAL INTEGRAL TRANSFORMS

The method we used to solve the one-wave propagation problem is totally transposable, and the only main originality here is to apply two different two-sided spatial Laplace transforms on the variable x whether φ or ψ is considered. The overbar and the tilde refer to φ and ψ, respectively. Hence, to be concise, many intermediary calculations and mathematical aspects of the presentation will be omitted. Thus, let us define the one-sided Laplace transforms with respect to time as

$$\bar{\varphi}(x,z,s)=\int_0^{+\infty}\varphi(x,z,t)\exp(-st)dt,$$

$$\bar{\psi}(x,z,s)=\int_0^{+\infty}\psi(x,z,t)\exp(-st)dt, \tag{7}$$

the double-sided Laplace transforms

$$\tilde{\bar{\varphi}}(p,z,s)=\int_{-\infty}^{+\infty}\bar{\varphi}(x,z,s)\exp(f(s)px)dx,$$

$$\hat{\bar{\psi}}(p,z,s)=\int_{-\infty}^{+\infty}\bar{\psi}(x,z,s)\exp(g(s)px)dx, \tag{8}$$

and use the following viscoelastic correspondences:

$$\mu_o\leftrightarrow s\bar{\lambda}(s), \quad \mu_o\leftrightarrow s\bar{\mu}(s), \tag{9}$$

where

$$\bar{\lambda}(s)=\frac{\lambda_o}{\lambda_o+\lambda_1}\left(\frac{\lambda_1}{s}+\frac{\lambda_o}{s+1/\tau_o}\right),$$

$$\bar{\mu}(s)=\frac{\mu_o}{\mu_o+\mu_1}\left(\frac{\mu_1}{s}+\frac{\mu_o}{s+1/\omega_o}\right), \tag{10}$$

are the time Laplace transforms of the relaxation functions, with

$$f(s)=\sqrt{s\frac{\lambda_o+2\mu_o}{\bar{\lambda}(s)+2\bar{\mu}(s)}}=c_d\sqrt{\frac{s\rho}{\bar{\lambda}(s)+2\bar{\mu}(s)}},$$

$$g(s)=c_s\sqrt{\frac{s\rho}{\bar{\mu}(s)}}, \tag{11}$$

then we seek for two unknown functions $\tilde{\bar{A}}(p,s)$ and $\tilde{\bar{B}}(p,s)$ such that

$$\tilde{\bar{\varphi}}(p,z,s)=\tilde{\bar{A}}(p,s)\exp(-f(s)\eta_d z),$$

$$\hat{\bar{\psi}}(p,z,s)=\hat{\bar{B}}(p,s)\exp(-g(s)\eta_s z), \tag{12}$$

$$\eta_d=\sqrt{1/c_d^2-p^2}, \quad \eta_s=\sqrt{1/c_s^2-p^2}. \tag{13}$$

Note that although at this step the dual variables s and p can be considered as real numbers, the functions f, g, η_d, and η_s can be complex valued; thus, in order to both verify Eq. (6) and uniquely define the square root function, we choose the branch cut along the negative real axis of the complex plane, for instance, $]-\infty,-1/c_d]$ and $]-\infty,-1/c_s]$ for η_d and η_s, respectively, in the complex p-plane. The real and imaginary parts of complex number will be noted Im and Re, respectively.

Now, because $\tilde{\bar{A}}(p,s)$ and $\hat{\bar{B}}(p,s)$ must verify the boundary-initial conditions (4) and (5), we obtain the following system:

$$\begin{pmatrix}\left(\frac{\bar{\lambda}(s)+2\bar{\mu}(s)}{c_d^2\bar{\mu}(s)}-2p^2\right)f(s) & 2p\eta_s g(s)\\ -2p\eta_d f(s) & \left(\frac{1}{c_s^2}-2p^2\right)g(s)\end{pmatrix}\begin{pmatrix}\tilde{\bar{A}}(p,s)\\ \hat{\bar{B}}(p,s)\end{pmatrix}=\begin{pmatrix}-\frac{F}{\bar{\mu}(s)}\\ 0\end{pmatrix}, \tag{14}$$

where the Dirac delta function is defined either from the first or the second following double equality:

$$\delta(x)=\frac{f(s)}{2i\pi}\int_{-i\infty+a}^{+i\infty+a}\exp(-f(s)px)dx=\frac{g(s)}{2i\pi}\int_{-i\infty+a}^{+i\infty+a}\exp(-g(s)px)dx, \tag{15}$$

depending on whether we extract $\tilde{\bar{A}}(p,s)$ or $\hat{\bar{B}}(p,s)$, respectively, in Eq. (14), where a is a positive real number, and i

denotes the unit complex number. The potentials can be written as

$$\varphi(x,z,t) = -\frac{F}{2i\pi}\int_{-i\infty+a}^{+i\infty+a}\frac{1}{\bar{\mu}(s)}\exp(st)ds\frac{1}{2i\pi}\int_{-i\infty+b}^{+i\infty+b}\frac{1/c_s^2-2p^2}{R(p,s)} \times\exp(-f(s)(\eta_d z+px))dp, \tag{16}$$

$$\psi(x,z,t) = -\frac{F}{2i\pi}\int_{-i\infty+a}^{+i\infty+a}\frac{1}{\bar{\mu}(s)}\exp(st)ds\frac{1}{2i\pi}\int_{-i\infty+b}^{+i\infty+b}\frac{2p\eta_d}{R(p,s)} \times\exp(-g(s)(\eta_s z+px))dp, \tag{17}$$

where b is a real number belonging to the strip of convergence defined by $-1/c_d<\mathrm{Re}(s)<1/c_d$, whereas κ and the Rayleigh function R write

$$\kappa(s) = \frac{\bar{\lambda}(s)+2\bar{\mu}(s)}{\bar{\mu}(s)} \quad \text{and} \quad R(p,s) = \left(\frac{\kappa(s)}{c_d^2}-2p^2\right)\left(\frac{1}{c_s^2}-2p^2\right) + 4\eta_d\eta_s p^2, \tag{18}$$

respectively. First, given that s is initially considered to be a positive real number, the inner integrals in Eqs. (16) and (17) are performed via the well-known Cagniard change in variable. Second, the outer integrals on s are performed in the complex s-plane along appropriate Bromwich's path by analytical continuation.[5] These two integrations are thoroughly described in our previous publication on one-wave propagation problem,[1] and the calculations are not repeated here. Particularly, it is expedient to show that the branch points,

$$s_0 = 0, \quad s_{a_1} = -\frac{1}{\tau_o}, \quad s_{a_2}, s_{a_3}, s_{b_1} = -\frac{1}{\omega_o}, \quad s_{b_2} = -\frac{\mu_o}{\mu_o+\mu_1}\frac{1}{\omega_o}, \tag{19}$$

are not essential singularities, which verify with, for instance, $\omega_o>\tau_o$,

$$s_{a_1}<s_{a_2}<s_{b_1}<s_{a_3}<s_{b_2}<s_0, \tag{20}$$

where s_{a_2} and s_{a_3} are the two negative real solutions of the following quadratic equation:

$$\bar{\lambda}(s)+2\bar{\mu}(s)=0. \tag{21}$$

Moreover the Rayleigh pole location on the complex p-plane depends on the function $\kappa(s)$ but is still located on the positive real axis as s is initially assumed to be a positive real number. In order to shorter the presentation, the displacement on the free surface for which the Rayleigh pole contribution is involved is not discussed here. We will deal only with an arbitrary point in the interior of the half space.

IV. PROBLEM SOLUTION

After the Cagniard changes in variables,[6]

$$\eta_d z+px=\tau \quad \text{and} \quad \eta_s z+px=\tau, \tag{22}$$

the inner integrals in Eqs. (16) and (17) are integrated along the Cagniard contours $p(\tau)$ parametrized with the positive real variable τ. This procedure is now well known for solving the corresponding elastic solution.[7,8] The outer integrals in Eqs. (16) and (17) are performed by analytical continuation around the branch cuts located on the real axis of the complex s-plane $[s_{a_1},s_{a_2}]\cup[s_{b_1},s_{a_3}]$ and $[s_{b_1},s_{b_2}]$, respectively. The final result for the interior points, $z>0$, repaired in polar coordinates (r,θ), is concisely summarized as follows:

$$\varphi(x,z,t) = H(t-t_d)\frac{F}{\pi^2}\int_{s\in I}\frac{1}{\bar{\mu}(s)}ds\int_{t_d}^{t}\mathrm{Im}\left[\alpha(p_d(\tau))\frac{dp_d(\tau)}{d\tau}\right] \times\cos(f(s)(t-\tau))d\tau, \tag{23}$$

$$\psi(x,z,t) = H(t-t_s)\frac{F}{\pi^2}\int_{s\in J}\frac{1}{\bar{\mu}(s)}ds\int_{t_s}^{t}\mathrm{Im}\left[\beta(p_s(\tau))\frac{dp_s(\tau)}{d\tau}\right] \times\cos(g(s)(t-\tau))d\tau + H(t-t_{ds})H(\theta-\theta_c)\frac{F}{\pi^2}\int_{s\in J}\frac{1}{\bar{\mu}(s)}ds\int_{t_{ds}}^{t}\mathrm{Im}\left[\beta(p_{ds}(\tau))\frac{dp_{ds}(\tau)}{d\tau}\right] \times\cos(g(s)(t-\tau))d\tau, \tag{24}$$

where

$$I=[s_{a_1}+i\varepsilon,s_{a_2}+i\varepsilon]\cup[s_{b_1}+i\varepsilon,s_{a_3}+i\varepsilon],$$

$$J=[s_{b_1}+i\varepsilon,s_{b_2}+i\varepsilon], \quad (\varepsilon\to 0^+), \tag{25}$$

$$\alpha(p)=\frac{1/c_s^2-2p^2}{R(p,s)}, \quad \beta(p)=\frac{2p\eta_d}{R(p,s)}, \tag{26}$$

$$p_d(\tau)=c_d\tau/r\,\sin\theta+i[(c_d\tau/r)^2-1]^{1/2}\cos\theta, \tag{27}$$

$$p_s(\tau)=c_d\tau/r\,\sin\theta+ik[(c_s\tau/r)^2-1]^{1/2}\cos\theta, \tag{28}$$

$$p_{ds}(\tau)=c_d\tau/r\,\sin\theta+ik[1-(c_s\tau/r)^2]^{1/2}\cos\theta, \quad (k=c_d/c_s), \tag{29}$$

$$r=\sqrt{x^2+z^2}, \quad \theta=\arctan(x/z), \quad \theta\in[0,\pi/2], \tag{30}$$

$$t_d=r/c_d, \quad t_s=r/c_s, \quad t_{ds}=r/c_d[\sin\theta+\sqrt{k^2-1}\cos\theta], \quad \theta_c=\arcsin(1/k), \tag{31}$$

where H denotes the unit step function, t_d, t_s, and t_{sd} denote the travel times of the dilatational, shear, and head waves, respectively. Note briefly that the second term in Eq. (24) identifies the head wave, as the contour $p_s(\tau)$ intercepts the branch cut on the real p-axis, $[1/c_d,+\infty$ (, for $\theta>\theta_c$.

V. CONCLUSION

A full resolution of the viscoelastic half space Lamb problem was proposed from an extension of Cagniard's method. More generally, the developed method of resolution can be the basis for natural extension from elastodynamics to viscoelastodynamics of numerous existing analytical problem solutions. This extension can help the acoustician to a

better understanding of the wave conversion at an interface in presence of dispersion via the generation of the head wave.

[1]A. Moura, "Causal analysis of transient viscoelastic wave propagation," J. Acoust. Soc. Am. **119**, 751–755 (2006).
[2]R. L. Ludwig and P. L. Levin, "Analytical and numerical treatment of pulsed wave propagation into a viscous fluid," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 789–792 (1995).
[3]E. Strick, "Propagation of elastic wave motion from an impulsive source along a fluid/solid interface. II. Theoretical pressure pulse," Philos. Trans. R. Soc. London, Ser. A **251**, 465–523 (1959).
[4]A. Ben-Menahem and S. J. Singh, *Seismic Waves and Sources* (Springer-Verlag, New York, 1981).
[5]D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, 1941).
[6]L. Cagniard, *Reflection and Refraction of Progressive Seismics Waves*, translated by E. A. Flinn and C. H. Dix (McGraw-Hill, New York, 1962).
[7]J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).
[8]K. Aki and P. G. Richards, *Quantitative Seismology* (University Science Books, Sausalito, CA, 2002).

Observation of energy cascade creating periodic shock waves in a resonator (L)

Tetsushi Biwa
Department of Mechanical Systems and Design, Tohoku University, Sendai 980-8579, Japan

Taichi Yazaki
Department of Physics, Aichi University of Education, Kariya 448-8542, Japan

(Received 31 August 2009; revised 15 December 2009; accepted 18 December 2009)

Nonlinear excitation of periodic shock waves in high-amplitude standing waves was studied from measurements of the acoustic intensity. A gas column of atmospheric air in a cylindrical resonator was driven sinusoidally by an oscillating piston at the fundamental resonance frequency. Acoustic pressure and axial acoustic particle velocity were simultaneously measured and decomposed into the Fourier components, from which the intensity associated with each of the oscillating modes in the resonator was determined. This letter reports the energy cascade from the driven mode to the second harmonic in the periodic shock waves in the resonator.
 [DOI: 10.1121/1.3291029]

PACS number(s): 43.25.Cb, 43.35.Ud [RR] Pages: 1189–1192

When a gas column in an acoustic resonator of finite Q value is periodically perturbed by a piston oscillating at frequencies away from the resonance frequencies, the gas oscillations are continuous. But as the driving frequency approaches the fundamental resonance frequency, nonlinear effects in gas dynamics give rise to periodic shock waves.[1–5] Notably in the long history of the study of this problem,[6] a great contribution has been made by Chester.[1] He derived a nonlinear differential-integral equation to solve the problem to second order from basic equations of hydrodynamics and succeeded in obtaining either continuous or discontinuous solutions depending on the frequency. The temporal waveform predicted by his solution was experimentally verified by Cruikshank[2] through measurements of the acoustic pressure at the rigid end of the resonator.

Wave distortion in the time domain is thought of as generation of harmonic oscillations in the frequency domain. The idea of energy cascade presented by Coppens and Sanders[7] explains the successive energy transfer from the driven mode to the higher harmonics in finite-amplitude standing waves, which led to development of the novel methods for suppressing shock waves. Lawrenson *et al.*[8] achieved a very high-amplitude shock-free oscillation using a *dissonant* resonator in which the overtones are not integer multiples of the fundamental frequency.[9] Sugimoto *et al.*[10,11] made use of wave dispersion caused by an array of Helmholtz resonators and successfully created a solitary sound wave instead of the shock wave. These methods play important roles in upgrading the performance of acoustic compressors and thermoacoustic heat engines since excitation of the harmonics limits the attainable acoustic pressure to about 10% of the filling pressure.[12] Although techniques have been found for inhibiting shock formation, little experimental attention has been directed to the question of how the acoustic energy is pumped up from the fundamental mode to the harmonics.

In order to answer this question, we focus on the acoustic intensity I of the gas column. Measurements of I have become possible recently through the simultaneous measurements of acoustic pressure P and axial acoustic particle velocity U (Ref. 13) and have been applied to the study of spontaneous gas oscillations in thermoacoustic systems.[14,15] The observation of the positive *work source* gave clear experimental insight into the generation of sound by heat.[16–19] We apply this method to the problem of nonlinear acoustic gas oscillation and observe how harmonics are generated. This paper documents the energy cascade which creates the shock wave in the resonator.

Acoustic intensity I represents the time-averaged acoustic power flux per unit cross-sectional area and is given by

$$I = \langle PV \rangle, \tag{1}$$

where V is the cross-sectional average of U, and angular brackets represent the time average. The *work source*

$$w = \mathrm{div}\, I \tag{2}$$

is equivalent to the time-averaged acoustic power production per unit volume. For a resonator closed at one end and driven by an acoustic driver at the other end, I flows down the resonator and its magnitude monotonically decreases to zero at the closed end. The negative w in this case means the rate of acoustic power dissipation, whereas a positive w represents acoustic power production.[16–19] Within the framework of linear acoustic theory, w is accounted for by surface attenuations at the tube wall due to viscosity and thermal conductivity of the gas and can be evaluated if the acoustic amplitude is given, as we will show later. We confirmed that the experimental w agreed with the theoretical one within the experimental error when the acoustic pressure was below 1% of the mean pressure.[20] Thus, a theoretical w given from the acoustic amplitude can serve as a good measure of acoustic power dissipation caused through a linear damping mechanism.

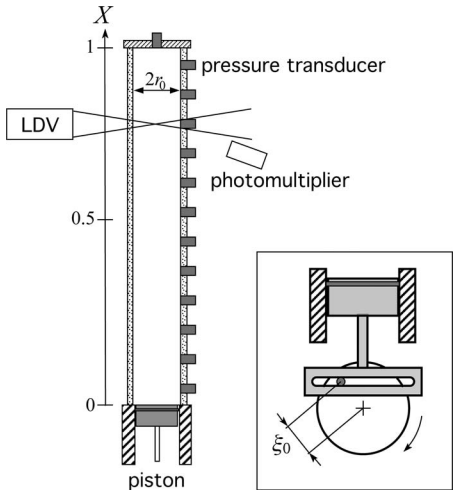


FIG. 1. Experimental setup. The oscillating piston at the bottom of the resonator was driven by a scotch yoke mechanism shown in the inset. Axial coordinate X was normalized with respect to the resonator length L. Pressure and velocity were measured at evenly spaced positions along X.

A schematic illustration of the experimental setup is shown in Fig. 1. Air at atmospheric pressure was confined in a cylindrical glass tube of length $L=1.15$ m and internal radius $r_0=10.5$ mm. One end of the tube was closed by a rigid plate, while the other was connected to a piston unit having a constant displacement amplitude $\xi_0=0.7$ mm. The unit employs a scotch yoke mechanism which ensures the piston's pure sinusoidal motion, as noted by Merkli and Thomann.[21] Thus, our acoustic driver excites only the fundamental oscillation and has no direct connection to the generation of harmonic oscillations. Before measurements of the acoustic intensity I, the shock region was confirmed from observation of the acoustic pressure at the closed end when the operating frequency of the piston was changed. It was found that the shock appeared in the region from 142 to 147 Hz, which coincided with the fundamental resonance frequency of the tube.

Simultaneous measurements of P and U were conducted at the frequency $f=144.4$ Hz. The corresponding piston's Mach number $M_P=\omega\xi_0/a$ is 1.8×10^{-3}, where ω $(=2\pi f)$ is the angular frequency of the piston, and a is the speed of sound in air (=344 m/s). The pressure $P(X,t)$ along the tube axis was measured by small pressure transducers (Kulite XT-190, natural frequency 350 kHz). The transducers were flush mounted on the tube wall and were arranged at an equidistant spacing from each other. The axial velocity $U(X,t)$ on the central axis of the tube was measured at the same axial positions as the pressure measurements using a laser Doppler velocimeter. Signal voltages of the pressure and velocity were sampled at high and low sampling frequencies of 102.4 and 2.56 kHz using a 24-bit spectrum analyzer (Onosokki Graduo).

Figure 2 shows the spatial and temporal evolutions of the periodic shock wave induced in the resonator. The time series of $P(X,t)$ and $U(X,t)$ sampled at the higher sampling frequency (102.4 kHz) is plotted with vertical offsets proportional to the measurement positions. It is seen that the shock front travels back and forth in the resonator at the speed of sound, bouncing off the solid end once per acoustic period $(=1/f)$. The pressure and velocity measured at the same axial position show the discontinuity at the same time, but their waveforms are significantly different. The pressure tends to form a sawtooth wave, whereas the velocity tends to form a rectangular wave with the highest peak to peak velocity at

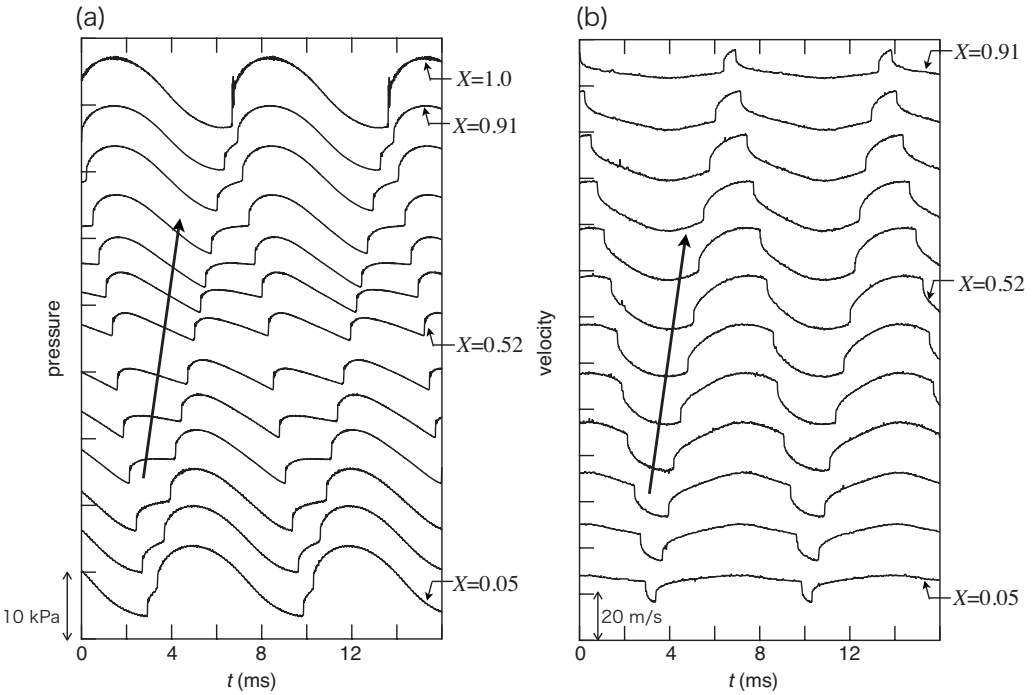


FIG. 2. Spatial and temporal variations in the periodic shock wave in the resonator. Pressure (a) and central velocity (b) are shown with vertical offsets according to their axial positions. Data with the largest X near the closed end are plotted at the top, whereas those near the driver end are plotted at the bottom. Arrows represent the propagation of the shock front.

the middle of the resonator corresponding to the pressure node.

Several experimental studies report the stability diagram of the oscillatory flow in tubes in terms of the Reynolds number $\mathrm{Re}_{\mathrm{OS}}$ based on the velocity amplitude defined as $\mathrm{Re}_{\mathrm{OS}} \equiv 2r_0 V/\nu$, and the square root of the non-dimensional frequency $\sqrt{\omega'}$ given by $\sqrt{\omega'} = \sqrt{2} r_0/\delta$, where δ is the viscous penetration depth ($\delta = \sqrt{2\nu/\omega}$, where ν is the kinematic viscosity of the gas). Acoustic turbulence takes place in the region satisfying both $2 \times 10^3 < \mathrm{Re}_{\mathrm{OS}}$ and $780\sqrt{\omega'} < \mathrm{Re}_{\mathrm{OS}}$.[21–23] The largest $\mathrm{Re}_{\mathrm{OS}}(=1.7 \times 10^4)$ in the present experiment ($\sqrt{\omega'} = 81$) was far below the critical Reynolds number $\mathrm{Re}_{\mathrm{OS}} \sim 6.3 \times 10^4$. Furthermore, we did not observe turbulent bursts which should appear as the manifestation of turbulent flow. These facts permit us to treat the present experimental data as laminar flow.

In order to determine I associated with each oscillating mode, we decompose the pressure and velocity sampled at the lower sampling frequency into Fourier components as

$$P = \sum_n p_n e^{in\omega t} \tag{3}$$

and

$$U = \sum_n u_n e^{in\omega t} \tag{4}$$

via the fast-Fourier-transform algorithm, where p_n and u_n are the complex amplitudes of the nth component of pressure and velocity. The fundamental and the second harmonic mode correspond to $n=1$ and $n=2$, respectively. The acoustic intensity I is rewritten as

$$I = \sum_n I_{2n}, \tag{5}$$

where

$$I_{2n} = \tfrac{1}{2}|p_n||v_n|\cos\,\phi_n. \tag{6}$$

Here, v_n represents the nth component of the cross-sectional mean velocity V, and $\phi_n(=\arg(p_n/v_n))$ is the phase lead of v_n to p_n. The velocity component v_n was determined from the measured central velocity u_n on the basis of the laminar oscillating flow theory.[20]

Figure 3(a) shows the experimental acoustic intensity I_2 determined using Eq. (6). The sign of I_2 was positive everywhere in the resonator, showing that the acoustic power flux consistently flowed in the direction from the driver end ($X=0$) to the closed end ($X=1$). This means that the solid piston supplies the acoustic power necessary to sustain the gas oscillation. The negative work source $w_2 = dI_2/dx$ ($x=LX$) observed throughout the resonator represents the dissipative power per unit volume and time, the average of which is found to be -1430 W/m^3. This is significantly larger than that expected in a linear standing wave with the same acoustic amplitude, as we show below.

Linear acoustic theory states that the total dissipation $\dot{E}$ is given by[12]

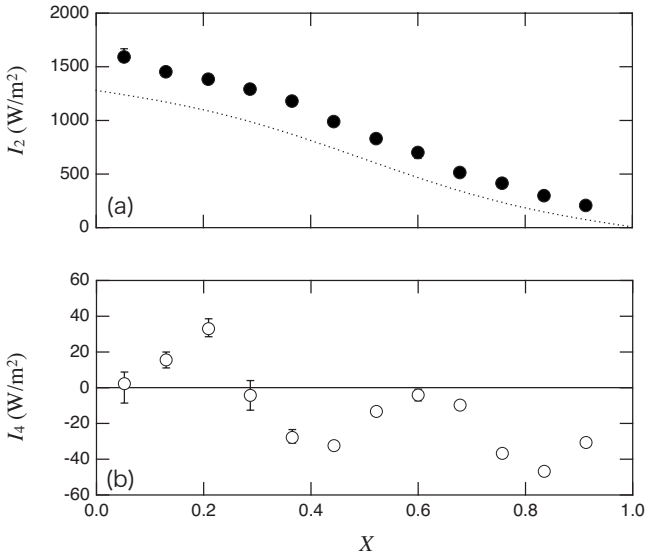


FIG. 3. Axial distribution of the acoustic intensity I_2 (a). The dotted curve represents the theoretical intensity estimated using the measured $p_{\max}$ through the linear acoustic theory. The intensity I_4 is shown in (b).

$$\dot{E} = \omega\pi r_0^2 L \frac{p_{\max} v_{\max}}{4a} \frac{\delta}{r_0}\left(1 + \frac{\gamma - 1}{\sqrt{\sigma}}\right), \tag{7}$$

where $p_{\max}$ and $v_{\max}$ are the maximum amplitudes of the pressure and velocity, and γ and σ are the ratio of isobaric to isochoric specific heats and Prandtl number of the gas. The average rate of the energy dissipation per unit volume, which we call the theoretical work source, is simply given by $\dot{E}$ divided by the internal volume $\pi r_0^2 L$. Using observed $p_{\max}$ (=5170 Pa) and $v_{\max}$ (=12.4 m/s), we obtain a theoretical work source of -1120 W/m^3 for the present acoustic oscillation. Therefore, the experimental work source w_2 exceeds the theoretical one by 28%. The theoretical work source should account for the surface attenuation due to the viscosity and the thermal conductivity of the gas. Thus, we attribute the extra damping given by the difference $\Delta w_2 = -310$ W/m^3 of the experimental work source w_2 from the theoretical one to the acoustic power consumed for the excitation of the higher harmonics.

Figure 3(b) shows the acoustic intensity I_4 associated with the second harmonic. Even though a positive work source is not found in the fundamental mode, it is clearly observed that I_4 increases in regions $0 < X < 0.2$, $0.4 < X < 0.6$, and $0.8 < X < 1.0$, demonstrating the acoustic power production. This positive w_4, reaching 170 W/m^3 at $X = 0.13$, represents the energy pumping from the driven mode to the second harmonic. Disappearance of I_4 at the driver end is reasonable because of the pure sinusoidal motion of the piston. Therefore, the second harmonic is produced by the gas column through nonlinear interaction between the oscillating modes. A very steep negative slope (-140 W/m^3) is observed near the sink of I_4 at $X=0.3$ that I_4 flows into from both sides, and the flow direction of I_4 becomes opposite when $X > 0.3$. The theoretical work source caused by the linear damping mechanism is estimated to be -91 W/m^3 using Eq. (7) based on the axial distribution of the Fourier components p_2 and v_2. The rest of Δw_2 should have a close

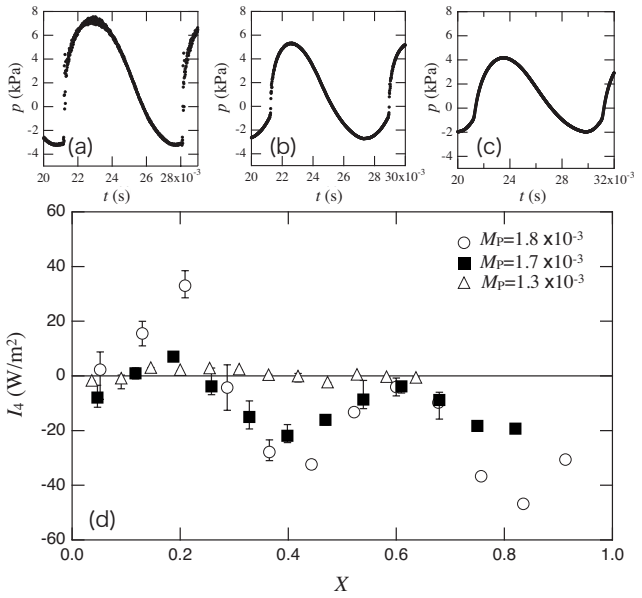


FIG. 4. Temporal variation in the pressure observed at the rigid end of the resonator when $M_P=1.8\times10^{-3}$ (a), 1.7×10^{-3} (b), and 1.3×10^{-3} (c). Axial distribution of the acoustic intensity I_4 is plotted in (d). Symbols correspond to the data with $M_P=1.8\times10^{-3}$ (○), 1.7×10^{3} (■), and 1.3×10^{-3} (△), respectively.

relation to the energy cascade process to the higher harmonics with $n>3$.

In order to see how the characteristic behavior of I_4 is related to the generation of the shock waves, we conducted the same measurements with the piston's Mach numbers $M_P=1.7\times10^{-3}$ and 1.3×10^{-3}. Smaller values of M_P were achieved by using longer glass cylinders, which reduces the fundamental resonance frequency. The piston displacement was kept at $\xi_0=0.7$ mm. Figures 4(a)–4(c) show the temporal pressure waveforms observed at the closed end, and Fig. 4(d) shows the corresponding axial distribution of I_4. We can verify that the energy cascade is executed in the same axial positions, but it becomes gradually weaker with the lower M_P. In response to such changes, the shock disappears when M_P is the lowest, as shown in Fig. 4(c). Hence, it is concluded that the energy pumping from the second harmonic leads to the formation of the shock wave in the resonator. If we can suppress the positive w_4 distributed in the resonator, a new method would be found to annihilate shock waves in the resonator.

Although we have focused on the intensities I_2 and I_4 in the present study, the higher order intensities would be of importance in shock waves that have shock fronts reaching 80% of the mean pressure.[5] Intensity measurements may also give some insight into the fascinating nonlinear processes in high-amplitude acoustics such as the acoustic soliton[11] and acoustic chaos.[24]

This study was financially supported by Industrial Technology Research Grant Program in 2008 from New Energy and Industrial Technology Development Organization (NEDO) of Japan.

[1]W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).
[2]D. B. Cruikshank, "Experimental investigation of finite-amplitude acoustic oscillations in a closed tube," J. Acoust. Soc. Am. **52**, 1024–1036 (1972).
[3]R. A. Saenger and G. E. Hudson, "Periodic shock waves in resonating gas columns," J. Acoust. Soc. Am. **32**, 961–970 (1960).
[4]S. Temkin, "Nonlinear gas oscillations in a resonant tube," Phys. Fluids **11**, 960–963 (1968).
[5]B. Sturtevant, "Nonlinear gas oscillations in pipes. Part 2. Experiment," J. Fluid Mech. **63**, 97–120 (1974).
[6]For example, see M. A. Ilgamov, R. G. Zaripov, R. G. Galiullin, and V. B. Repin, "Nonlinear oscillations of a gas in a tube," Appl. Mech. Rev. **49**, 137–154 (1996).
[7]A. B. Coppens and J. V. Sanders, "Finite-amplitude standing waves in rigid-walled tubes," J. Acoust. Soc. Am. **43**, 516–529 (1968).
[8]C. Lawrenson, B. Lipkens, T. Lucas, D. Perkins, and T. Doren, "Measurements of macrosonic standing waves in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).
[9]D. F. Gaitan and A. A. Atchley, "Finite amplitude standing waves in harmonic and anharmonic tubes," J. Acoust. Soc. Am. **93**, 2489–2495 (1993).
[10]N. Sugimoto, M. Masuda, T. Hashiguchi, and T. Doi, "Annihilation of shocks in forced oscillations of an air column in a closed tube," J. Acoust. Soc. Am. **110**, 2263–2266 (2001); M. Masuda and N. Sugimoto, "Experiments of high-amplitude and shock-free oscillations of air columns in a tube with array of Helmholtz resonators," *ibid.* **118**, 113–123 (2005).
[11]N. Sugimoto, M. Masuda, J. Ohno, and D. Motoi, "Experimental demonstration of generation and propagation of acoustic solitary waves in an air-filled tube," Phys. Rev. Lett. **83**, 4053–4056 (1999); A. Hellemans, "Conjugating a solitary sound wave," Science **286**, 2062 (1999).
[12]G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[13]T. Yazaki and A. Tominaga, "Measurement of sound generation in thermoacoustic oscillations," Proc. R. Soc. London, Ser. A **454**, 2113–2122 (1998).
[14]A. Tominaga, "Thermodynamic aspects of thermoacoustic theory," Cryogenics **35**, 427–440 (1995); *Fundamental Thermoacoustics* (Uchida Roukakuho, Tokyo, Japan, 1998).
[15]G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Sewickley, PA, 2002).
[16]T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).
[17]T. Yazaki, T. Biwa, and A. Tominaga, "A pistonless Stirling cooler," Appl. Phys. Lett. **80**, 157–159 (2002).
[18]Y. Ueda, T. Biwa, U. Mizutani, and T. Yazaki, "Acoustic field in a thermoacoustic Stirling engine having a looped tube and resonator," Appl. Phys. Lett. **81**, 5252–5254 (2002); "Experimental studies of a thermoacoustic Stirling prime mover and its application to a cooler," J. Acoust. Soc. Am. **115**, 1134–1141 (2004).
[19]T. Biwa, Y. Tashiro, U. Mizutani, M. Kozuka, and T. Yazaki, "Experimental demonstration of thermoacoustic energy conversion in a resonator," Phys. Rev. E **69**, 066304 (2004).
[20]T. Biwa, Y. Ueda, H. Nomura, U. Mizutani, and T. Yazaki, "Measurement of the Q value of an acoustic resonator," Phys. Rev. E **72**, 026601 (2005).
[21]P. Merkli and H. Thomann, "Transition to turbulence in oscillating pipe flow," J. Fluid Mech. **68**, 567–575 (1975).
[22]M. Hino, M. Sawamoto, and S. Takasu, "Experiments on transition to turbulence in an oscillatory pipe flow," J. Fluid Mech. **75**, 193–207 (1976).
[23]M. Ohmi and M. Iguchi, "Critical Reynolds number in an oscillating pipe flow," Bull. JSME **25**, 165–172 (1982).
[24]T. Yazaki, S. Takashima, and F. Mizutani, "Complex quasiperiodic and chaotic states observed in thermally induced oscillations of gas columns," Phys. Rev. Lett. **58**, 1108–1111 (1987); T. Yazaki, S. Sugioka, F. Mizutani, and H. Mamada, "Nonlinear dynamics of a forced thermoacoustic oscillation," *ibid.* **64**, 2515–2518 (1990); T. Yazaki, "Experimental observation of thermoacoustic turbulence and universal properties at the quasiperiodic transition to chaos," Phys. Rev. E **48**, 1806–1818 (1993).

Active control of one-dimension impulsive reflection based on a prediction method (L)

Ning Han,[a] Shengzhen Feng, and Xiaojun Qiu
Key Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 6 March 2009; revised 19 July 2009; accepted 31 December 2009)

An approach for predicting the reflected sound pressure is proposed in one-dimensional sound field. For a duct ended with a rigid reflective surface, only one microphone is required to measure the total sound pressure on the surface, which is further used as the error sensing strategy in an active noise control (ANC) system to reduce the in-duct reflection. Experiments are carried out to validate the prediction method, and a broadband feedforward ANC system is implemented to suppress the impulsive reflection. It is found that the ANC system based on the reflected sound prediction is effective, and 12.2 dB attenuation of the one-dimension impulsive reflection is obtained after the active control. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3295691]

PACS number(s): 43.50.Ki [ADP] Pages: 1193–1196

I. INTRODUCTION

Active noise control (ANC) usually aims at reducing the total sound pressure radiated by a noise source; it may also be applied to reduce the reflected sound from a reflective body, which is useful for extending the low frequency range of wedges for building anechoic rooms or making an object invisible to incident detecting waves. Bao *et al.*[1–3] proposed a piezocomposite actuator for absorbing both the reflection and the transmission of normally incident waves. Experiments were presented for one-dimensional incident plane waves. Lafleur *et al.*[4,5] used the piezoelectric material to form active surfaces. Measurements in a tube indicated that the single layer can eliminate either the reflection or transmission, while the double layers eliminated both of them.

Guicking and Karcher[6] investigated the active impedance control for one-dimensional sound. The reflection coefficient was smaller than 0.05 from 100 to 1000 Hz. Orduña-Bustamante and Nelson[7] made a secondary source to be an active sound absorber in a duct, which can work with periodic, random, and transient input signals. Smith *et al.*[8] implemented a broadband hybrid absorption system in an impedance tube and found that the impedance matching method can provide better absorption than that achieved in the pressure-release method. Finneran and Hastings[9] adjusted the acoustic pressure/particle velocity ratio to match that of a plane wave in an active control system. Recently, Cobo and Cuesta[10,11] confirmed that releasing the sound pressure at the rear face of materials can provide active absorption at normal and oblique incidences. The direct and reflected events were picked up by time windowing.

For the direct control of the in-duct reflection, the wave separation algorithm[12] was used to separate the reflected and the incident waves, where the reflected sound pressure at the midpoint of two microphones is calculated through the measured signals. The two microphones cannot be positioned practically too close for the low frequency limitation or too far for the high frequency limitation. Friot and Bordier[13] introduced a method to distinguish the three-dimension scattered sound pressure from the total sound field and then suppressed it by ANC system. Preliminary result was given for one-dimension cases with more than 10 dB reflected noise reduction below 800 Hz. But the method for distinguishing the reflection is complex to implement, and the experimental control effect can only be obtained by comparing the noise without and with the rigid reflective surface. In the present paper, the impulsive wave will be chosen as the primary sound; thus the control effect can be obtained directly by comparing the amplitudes of the impulsive reflections before and after the active control.

Virtual sensing is a novel active control technique, and microphone arrays were used to predict the information at the virtual error sensor.[14–16] The concept of the virtual error sensing will be employed in this paper differently since the reflected sound pressure, which needs to be controlled, cannot be obtained directly by locating a physical sensor there. In the present research, a method is proposed for predicting the in-duct reflected sound pressure, and it is further used in the error sensing strategy of an ANC system to reduce the reflected sound. Different from the pressure-release or the impedance matching method in the previous reflection control, where the attentions were paid on the sound pressure or the particle velocity in the total sound field, the proposed ANC system attenuates the in-duct reflected wave directly by predicting it, and for a rigid reflective surface, only one microphone is required. The simple and compact structure of the ANC system can be further extended into three-dimension cases and applied to create cost effective anechoic tiles.

II. BASIC THEORY

A. Prediction of the reflected sound in a one-dimension duct

For simplicity, the origin of the coordinate system is positioned on the reflective surface, shown in Fig. 1. The

[a]Author to whom correspondence should be addressed. Electronic mail: hanning@nju.edu.cn

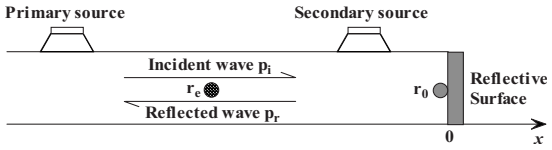


FIG. 1. ANC setup for the in-duct reflection. p_i the incident sound, p_r the reflected sound, r_e the location of the virtual error sensor, and r_0 a point on the reflective surface.

incident plane wave p_i with amplitude p_{ai} and the reflected plane wave p_r with amplitude p_{ar} are written as

$$p_i = p_{ai}e^{j(\omega t-kx)}, \quad p_r = p_{ar}e^{j(\omega t+kx)}, \tag{1}$$

where ω is the angle frequency, k is the wave number, and x indicates the position of the interested point on x-axis. Thus the total sound pressure and particle velocity are

$$p = (p_{ai}e^{-jkx} + p_{ar}e^{jkx})e^{j\omega t},$$

$$v = \left(\frac{p_{ai}}{\rho_0 c_0}e^{-jkx} - \frac{p_{ar}}{\rho_0 c_0}e^{jkx}\right)e^{j\omega t}, \tag{2}$$

where ρ_0 is the air density and c_0 is the sound velocity. On the reflective surface $x=0$, the sound pressure and the particle velocity are

$$p_0 = (p_{ai} + p_{ar})e^{j\omega t}, \quad v_0 = \left(\frac{p_{ai}}{\rho_0 c_0} - \frac{p_{ar}}{\rho_0 c_0}\right)e^{j\omega t}. \tag{3}$$

So p_{ai} and p_{ar} can be expressed as

$$p_{ai} = \tfrac{1}{2}(p_0 + \rho_0 c_0 v_0)e^{-j\omega t}, \quad p_{ar} = \tfrac{1}{2}(p_0 - \rho_0 c_0 v_0)e^{-j\omega t}. \tag{4}$$

In that way, the reflected sound pressure p_r can be predicted through the measured sound pressure and particle velocity on the reflective surface, i.e., $p_r=(p_0-\rho_0 c_0 v_0)e^{jkx}/2$, where e^{jkx} is regarded as the time delay $\tau=-x/c_0$. If the acoustic impedance ratio Z_s at $x=0$ is known, v_0 can be calculated by $v_0=p_0/Z_s$. For situations where the sound pressure is attenuated along with the propagation path, the wave attenuation should be added into the prediction equation $p_r=(p_0-\rho_0 c_0 v_0)e^{-\eta|x|}e^{jkx}/2$, and η is a positive number denoting the attenuation coefficient. For a rigid reflective surface, the particle velocity $v_0=0$, so the reflected sound pressure is predicted as $p_r=p_0 e^{-\eta|x|}e^{jkx}/2$.

B. ANC system for the reflected sound

In a duct ended with a rigid reflective surface, the ANC setup for the reflected sound is shown in Fig. 1. A microphone is positioned on the reflective surface r_0 for measuring the total sound pressure there, which is used to predict the reflected sound pressure at the virtual error sensor r_e. The secondary source is located between r_0 and r_e. The error signal in the ANC system for the in-duct reflection is comprised of the reflected sound pressure p_{pr} from the primary source and the total sound pressure $p_{cr}+p_{ci}$ from the secondary source

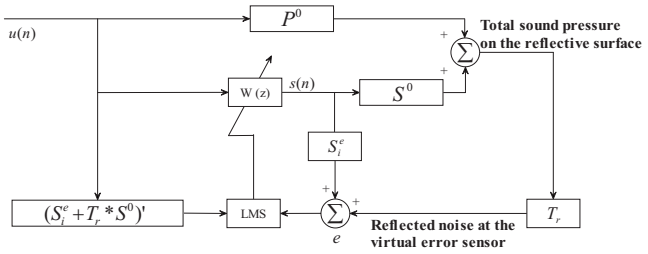


FIG. 2. Block diagram of the feedforward ANC system for the reflected sound. $W(z)$ the adaptive filter, $s(n)$ the input signal to the secondary source, and $'$ denotes the estimated FIR filter.

$$e = p_{pr} + p_{cr} + p_{ci}, \tag{5}$$

where $p_{pr}+p_{cr}=p_r=T_r*p_0$, p_0 is the measured sound pressure at r_0, $*$ is the time convolution, T_r denotes $e^{-\eta|x|}\delta(t-\tau)/2$ which is only applied for a rigid reflective surface, and $\delta(t)=1$ when $t=0$, otherwise $\delta(t)=0$. Defining S_i^e to be the impulse response from the secondary source input q_c to the direct secondary sound pressure at r_e, thus $p_{ci}=S_i^e*q_c$. The error signal e can be expressed by the sound pressure on the reflective surface p_0 and the secondary source input q_c,

$$e = T_r * p_0 + S_i^e * q_c. \tag{6}$$

Let S^0 be the impulse response from the secondary source input q_c to the total secondary sound pressure at r_0, and P^0 be the impulse response from the primary source input q_p to the total primary sound pressure at r_0. The ANC system for the reflected sound is shown in Fig. 2. Compared with the conventional feedforward ANC system with Filter-x LMS (FXLMS) algorithm, T_r*P^0 is regarded as the impulse response of the primary path, and $S_i^e+T_r*S^0$ is the impulse response of the secondary path. The measured total sound pressure p_0 on the reflective surface is used to predict the reflected sound pressure at the virtual error sensor r_e through the prediction filter T_r. With the identified S_i^e before control, the error signal can be obtained without using physical transducers at r_e.

III. EXPERIMENTS

Shown in Fig. 3, a 3.8 m long duct with cross-section dimensions of 0.165×0.166 m^2 (the cut-off frequency 1030 Hz) is employed in the experiment. An ADSP-21161N EZKIT digital signal processor (DSP) board is used as the signal generator for the loudspeaker input and the receiver of the microphone output, and further used as the controller in the ANC system with the sampling rate $f_s=4.8$ kHz.

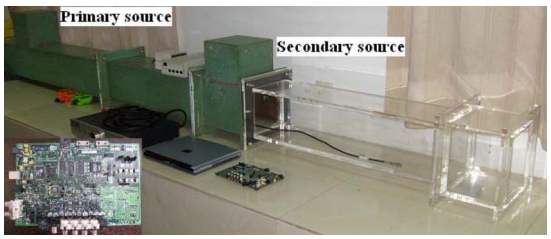
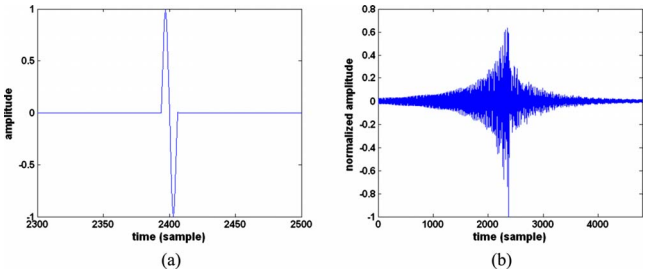


FIG. 3. (Color online) The experimental duct and the DSP board.



(a) (b)

FIG. 4. (Color online) (a) The desired observed signals and (b) the compensated input signals to the loudspeaker.

For clear observation in the experiments, the desired shape of the observed signals is shown in Fig. 4(a), where the impulsive part is a period of a sinusoidal wave at 400 Hz. The duration of the impulsive signal is 1 s, and the impulsive part is 0.0025 s. Its sound energy is mainly concentrated below 800 Hz, lower than the duct cut-off frequency; thus most of the sound waves are plane waves. In order to obtain the required waveform at the microphone, compensations are made for the loudspeaker input signal. The method is described as follows.

(1) Measuring the transfer function from the loudspeaker to the microphone

$$C(f) = \mathrm{FFT}\{c(n)\}, \tag{7}$$

where f denotes the frequency, $c(n)$ is the impulse response from the loudspeaker to the microphone, and n is the discrete time.

(2) Calculating the inverse transfer function from the loudspeaker to the microphone. Since zeros might be existent in the transfer function, a constant l is added to avoid the singular point in the inverse transfer function[17]

$$H(f) = \frac{C^*(f)}{C(f)C^*(f) + l}. \tag{8}$$

(3) Compensating the loudspeaker input signal. The needed output signal $y(n)$ is transferred into $Y(f)$ by fast Fourier transformation (FFT); thus the loudspeaker input is $Y_c(f) = Y(f)H(f)$. The driving signal to the loudspeaker is finally obtained by using the inverse FFT transformation

$$y_c(n) = \mathrm{IFFT}\{Y_c(f)\}. \tag{9}$$

The compensated input signals are shown in Fig. 4(b). By inputting the signals to the loudspeaker, a pulse in Fig. 4(a) can be observed at the microphone.

A. Validation of the prediction method for the reflected sound

The signals in Fig. 4(b) are input to the secondary source in Fig. 5(a). The signals measured at positions A and B are shown in the upper figure of Fig. 6(a). The ratio of the impulsive amplitudes at A and B is 0.9370, which is mainly due to the viscous losses at the surface of the walls as well as the structural acoustic coupling. The attenuation coefficient η is

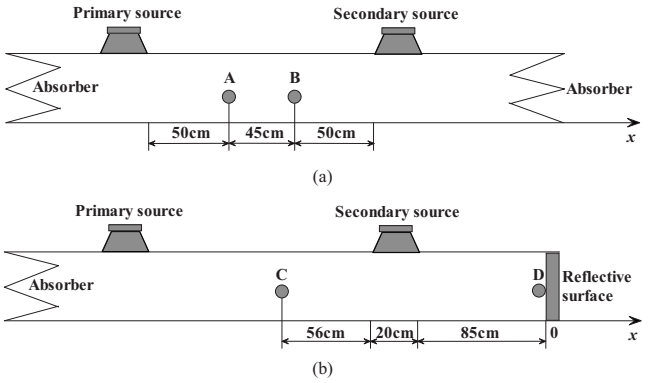


FIG. 5. The setup for obtaining (a) the attenuation coefficient in the duct and (b) the prediction equation of the reflected sound pressure directly.

calculated to be 0.1446; thus in the duct ended with a rigid reflective surface, the prediction equation for the reflected sound pressure at position C in Fig. 5(b) is

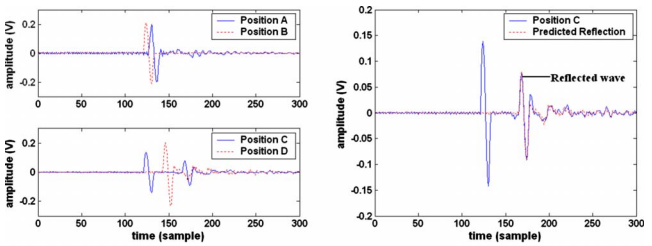
$$p_r = 0.3962 p_0 e^{jkx}. \tag{10}$$

The signals in Fig. 4(b) are input to the primary source in Fig. 5(b). The signals measured at positions C and D are shown in the lower figure of Fig. 6(a), and Fig. 6(b) shows the comparison between the reflected waves measured at C and predicted by Eq. (10). The predicted reflection is almost superposed on the measured one, expected for a little amplitude difference.

Comparing the amplitudes of the reflected sound measured at C and the total sound measured at D, the prediction equation for the reflected sound pressure at position C can be adjusted to be

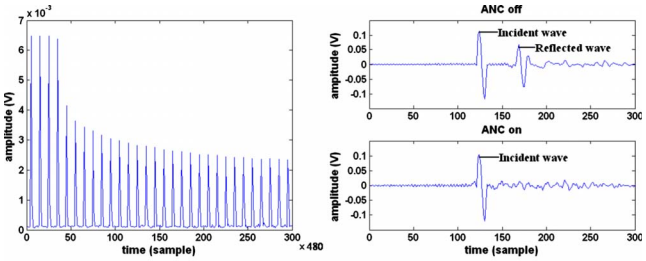
$$p_r = 0.3814 p_0 e^{jkx}, \tag{11}$$

whose amplitude is smaller than Eq. (10). This is mainly due to the un-uniform sound attenuation in the duct. For the experimental duct in Fig. 3, the sound attenuation in the first segment made of Perspex is different from that in the following wooden part; the secondary source is fixed on the duct by a wooden box, which might attenuate much sound and induce oscillations in the reflected wave. Also, the sound attenuation model $e^{-\eta|x|}$ may have deflections from the actual situation, which will induce errors in Eq. (10). In practice, special treatments should be implemented on the part of the secondary source in order to reduce its effect on the in-duct sound field. For the un-uniform inner surface, the duct attenuation $e^{-\eta|x|}$ should be obtained directly by comparing the



(a) (b)

FIG. 6. (Color online) (a) The measured signals in Fig. 5 and (b) the predicted reflection by Eq. (10) is compared with the measured reflection at C.



(a) (b)

FIG. 7. (Color online) (a) The attenuation process of the impulsive reflection; (b) the impulsive reflections in the duct before and after control.

measured amplitudes of the reflected sound at C and the total sound on the reflective surface D. Thus in the ANC experiment below, Eq. (11) is chosen as the prediction equation for the reflected sound pressure.

B. Suppressing the reflected sound by ANC system

In the experimental setup in Fig. 1, the duct is ended with a rigid reflective surface, and the virtual error sensor is located at position C in Fig. 5(b). According to the proposed ANC system in Fig. 2(a), where finite impulse response (FIR) filters with 256 taps are applied in the controller, the impulse response S_i^e from the secondary source input to the direct secondary sound pressure at C is measured when the reflective surface is absent, and the impulse response S^0 from the secondary source input to the total secondary sound pressure at the reflective surface is measured when the reflective surface is existent. In the active control for the in-duct reflection, a microphone is placed at C for observation. With the FXLMS algorithm in the ANC system, the error signal $e=T_r*p_0+S_i^e*q_c$ in Fig. 2(a) is reduced gradually to a steady level, and the reduction process is shown in Fig. 7(a), which indicates the curve of the root mean square for every 480 samples. The observed impulsive reflections without and with control are shown in Fig. 7(b), where 12.2 dB reduction is obtained after the active control.

For the non-rigid reflective surface, the particle velocity on the reflective surface should also be measured or calculated to predict the reflection for more general control conditions. Moreover, in the ANC system, the impulse response S_i^e from the secondary source to the direct secondary noise at the virtual error sensor is measured when the reflective surface is absent, which is not feasible in special situations where the reflective surface is not movable. This problem can be solved by measuring the impulse responses S^e and S^0 from the secondary source to the total secondary sound pressure at r_e and r_0 when the reflective surface is existent, and then the impulse response S_i^e is obtained approximately by $S_i^e=S^e-T_r*S^0$.

IV. CONCLUSIONS

Based on the sound field analysis in one-dimension duct, an approach is proposed for predicting the reflected sound pressure. For a rigid reflective surface, only one microphone is employed in the ANC system to form the error sensing strategy. The ANC system with the prediction method is validated to be effective, and 12.2 dB attenuation of the impulsive reflected sound is obtained after the active control. The introduced ANC system for the reflected sound is simple and compact and can be extended to the three-dimension cases, which will be more practical for building anechoic rooms with low cut-off frequency or making a reflective body invisible to the incident detecting waves.

The future work will mainly aim at the reflection control for the duct ended with a non-rigid reflective surface. Furthermore, the prediction method for the sound pressure scattered by three-dimension objects should be proposed in order to produce a compact ANC multi-channel system to control the scattered noise.

ACKNOWLEDGMENTS

Project No. 10674068 was supported by NSFC and NCET. We thank Mr. Xiaofeng Shi and Mr. Dongwei Shi for many helpful discussions.

[1]X. Bao, V. K. Varadan, V. V. Varadan, and T. R. Howarth, "Model of a bilaminar actuator for active acoustic control systems," J. Acoust. Soc. Am. **87**, 1350–1352 (1990).
[2]T. R. Howarth, V. K. Varadan, X. Bao, and V. V. Varadan, "Piezocomposite coating for active underwater sound reduction," J. Acoust. Soc. Am. **91**, 823–831 (1992).
[3]T. R. Howarth, X. Bao, R. Moser, V. K. Varadan, and V. V. Varadan, "Digital time delay network for an active underwater acoustic coating," J. Acoust. Soc. Am. **93**, 1613–1619 (1993).
[4]L. D. Lafleur, F. D. Shields, and J. E. Hendrix, "Acoustically active surfaces using piezorubber," J. Acoust. Soc. Am. **90**, 1230–1237 (1991).
[5]F. D. Shields and L. D. Lafleur, "Smart acoustically active surfaces," J. Acoust. Soc. Am. **102**, 1559–1566 (1997).
[6]D. Guicking and K. Karcher, "Active impedance control for one-dimensional sound," J. Vibr. Acoust. **106**, 393–396 (1984).
[7]F. Orduña-Bustamante and P. A. Nelson, "An adaptive controller for the active absorption of sound," J. Acoust. Soc. Am. **91**, 2740–2747 (1992).
[8]J. P. Smith, B. D. Johnson, and R. A. Burdisso, "A broadband passive–active sound absorption system," J. Acoust. Soc. Am. **106**, 2646–2652 (1999).
[9]J. J. Finneran and M. C. Hastings, "Active impedance control within a cylindrical waveguide for generation of low-frequency, underwater plane traveling waves," J. Acoust. Soc. Am. **105**, 3035–3043 (1999).
[10]P. Cobo and M. Cuesta, "Hybrid passive-active absorption of a microperforated panel in free field conditions," J. Acoust. Soc. Am. **121**, EL251–EL255 (2007).
[11]P. Cobo and M. Cuesta, "Measuring hybrid passive-active sound absorption of a microperforated liner at oblique incidence," J. Acoust. Soc. Am. **125**, 185–190 (2009).
[12]H. Zhu, R. Rajamani, and K. A. Stelson, "Active control of acoustic reflection, absorption, and transmission using thin panel speakers," J. Acoust. Soc. Am. **113**, 852–870 (2003).
[13]E. Friot and C. Bordier, "Real-time active suppression of scattered acoustic radiation," J. Sound Vib. **278**, 563–580 (2004).
[14]J. M. Munn, B. S. Cazzolato, C. D. Kestell, and C. H. Hansen, "Virtual error sensing for active noise control in a one-dimensional waveguide: Performance prediction versus measurement," J. Acoust. Soc. Am. **113**, 35–38 (2003).
[15]D. Petersen, A. C. Zander, B. S. Cazzolato, and C. H. Hansen, "Optimal virtual sensing for active noise control in a rigid-walled acoustic duct," J. Acoust. Soc. Am. **118**, 3086–3093 (2005).
[16]C. D. Petersen, A. C. Zander, B. S. Cazzolato, and C. H. Hansen, "A moving zone of quiet for narrowband noise in a one-dimensional duct using virtual sensing," J. Acoust. Soc. Am. **121**, 1459–1470 (2007).
[17]I. Marshall, "The production of acoustic impulses in air," Meas. Sci. Technol. **1**, 413–418 (1990).

High frequency poroelastic waves in hydrogels

Piero Chiarelli
CNR Institute of Clinical Physiology, via Moruzzi 1, 56124 Pisa, Italy

Antonio Lanatà and Marina Carbone
Interdepartmental Research Center "E. Piaggio," Faculty of Engineering, University of Pisa, via Diotisalvi 2, 56126 Pisa, Italy

Claudio Domenici
CNR Institute of Clinical Physiology, via Moruzzi 1, 56124 Pisa, Italy

(Received 3 December 2008; revised 25 September 2009; accepted 22 December 2009)

In this work a continuum model for high frequency poroelastic longitudinal waves in hydrogels is presented. A viscoelastic force describing the interaction between the polymer network and the bounded water present in such materials is introduced. The model is tested by means of ultrasound wave speed and attenuation measurements in polyvinylalcohol hydrogel samples. The theory and experiments show that ultrasound attenuation decreases linearly with the increase in the water volume fraction β of the hydrogel. The introduction of the viscoelastic force between the bounded water and the polymer network leads to a bi-phasic theory, showing an ultrasonic fast wave attenuation that can vary as a function of the frequency with a non-integer exponent in agreement with the experimental data in literature. When β tends to 1 (100% of interstitial water) due to the presence of bounded water in the hydrogel, the ultrasound phase velocity acquires higher value than that of pure water. The ultrasound speed gap at $\beta=1$ is confirmed by the experimental results, showing that it increases in less cross-linked gel samples which own a higher concentration of bounded water. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3293000]

PACS number(s): 43.20.Jr, 43.20.Bi, 43.80.Cs, 43.80.Ev [ADP] Pages: 1197–1207

I. INTRODUCTION

The present work is motivated in defining a reliable model for ultrasound (US) wave propagation both in hydrogels[1–5] and in extra-cellular matrices of natural soft tissues.[6–8]

In fact, even if the structure of a synthetic hydrogel is somehow different from that one of extra-cellular matrices with proteoglycans exists a strong analogy between the macroscopic response of charged hydrogels and that one of natural matrices, such as that one of derma,[9] in the diffusional wave limit.

On this basis, the US model for hydrogels can be a theoretical tool for the study and the characterization of tissue-mimicking phantom for US thermal therapy and for the development of non-invasive assessment method of soft tissue stiffness.[10–12]

This possibility is confirmed by the current research investigations that clearly show how the knowledge of the link between the poroelastic characteristics of a biological tissue and its acoustical behavior is a source of information that can be used for non-invasive investigation techniques.[13,14]

On the other hand, the US propagation in natural hydrogels, mostly composed of water, is usually modeled by means of the wave equation that holds for liquids.[5] Even if this approach is sufficiently satisfying, it does not completely explain the experimental behavior of US.

There are many discrepancies between US propagation in water and in natural hydrogels of soft tissues both for transverse and longitudinal acoustic waves.[6–8]

As far as longitudinal waves in biological gels are concerned, in the US absorption frequency law $\nu^{(1+\delta)}$ (where ν is the US frequency) δ ranges between $\frac{1}{4}$ and $\frac{1}{2}$, while it holds that $\delta=1$ for water.[15]

The fractional value of δ cannot also be explained by means of the established bi-phasic theories that lead to a frequency dependence equal to that of water with $\delta=1$.[16,17]

The bi-phasic models of the acoustic wave propagation were born in geology and engineering sciences[18–22] from where they received a clear imprinting. All proposed theories[16–24] do not abandon the scheme of a solid matrix permeated by a well distinguishable fluid.

Actually, it is quite evident that gels constitute a possible exception where the solid phase and the liquid one are chemically interacting and the boundary is not simply defined. As a matter of fact, the presence of the bounded water around the polymer chains creates a bearing whose thickness and viscous force do not follow the fluid dynamics laws.[19,20]

In order to solve this theoretical leak, in the present paper, a viscoelastic matrix-fluid interaction specific for hydrogels is introduced.

The hydrogel structure is somewhere between a solid and a liquid. It consists of polymers, or long chain molecules, which are cross-linked to create an entangled network immersed in a liquid medium that fills the intra-molecular interstices.

The properties of gels are strongly influenced by the interaction of these two components.

First, gels were conceptualized as porous media consisting of two interpenetrating macroscopic substances (an elas-

tic and porous solid matrix and a fluid). An elegant and satisfying theory based on this approach is the poroelastic model developed by Biot.[16,17,23,24]

One of the main characteristics of the Biot theory is that the dissipative relaxation is mainly ascribed to the relative motion of fluid against solid, while dissipations within fluid and solid are neglected.

At low frequencies, the friction of the fluid that moves against the solid network leads to a "diffusional" wave, giving theoretical forecasts that well agree with many experimental results.[9,25–29] Instead, this solid-fluid interaction scheme becomes inadequate at high frequencies.

In order to obtain a more satisfying model to describe how ultrasounds propagate in hydrogels, we propose a new mechanism of interaction between the fluid (water) and the solid matrix (polymer network).

The main mathematical statement, introduced here, takes into account the presence of the bounded water attached to the polymeric network of the hydrogel. The fluid-matrix interaction is modeled by ascribing a viscoelastic force between the polymer matrix and the bounded water around it, while a viscous force between the "polymer-bounded water aggregate" and the interstitial free water is assumed.

This assumption leads to relevant outcomes concerning the US propagation in gels. As confirmed by the experiments, the effect of the bounded water in hydrogels on the US speed is quite relevant as well as the frequency dependence of their attenuation. The model of the US propagation in hydrogels can lead to the understanding of their behavior and how it takes origin from the poroelastic structure. The possibility to describe the US propagation in terms of the poroelastic characteristics of both cells and extra-cellular matrix may lead to US non-invasive technique for the evaluation of the tissue physiology since the permeability and elasticity of the cell structures are directly related to their state and functionality.

In the following we derive the phase velocity of longitudinal US plane waves and their attenuation in hydrogels (Sec. II) by using the proposed fluid-matrix interaction.

Then, in Sec. III, we compare the theoretical forecasts with the outcomes of experimental measurements.

II. POROELASTIC WAVES IN HYDROGELS

Biot's[16,17] theory is carried out by assuming that the fluid motion in the pores follows the Poiseuille flow. In this case, the characteristic boundary, where the velocity attains a certain percentage of its maximum, is known as the viscous skin depth[19,20] $(\eta_d/\pi\nu\rho_f)^{1/2}$, where ρ_f and η_d are the fluid mass density and its dynamic viscosity, respectively. In the low frequency limit, this layer becomes larger than the pore diameter and the velocity profile can be assumed parabolic. At higher frequencies, when the viscous skin depth is smaller than the pore size, Biot[17] added a compensating factor $F_{(\nu)}$ to the fluid-solid friction coefficient f to take into account for the divergence by the Poiseuille flow. He stated that $F_{(\nu)}$ increases according to the law $\lim_{\nu/\nu c\to\infty} F_{(\nu)} \propto (\nu/\nu_c)^{1/2}$, where $\nu_c = f/\rho_f$ and f is the zero frequency friction coefficient[23] (i.e., the inverse of the hydraulic permeability of the medium).

Given that pore dimensions are very small in gels, the crossover frequency ν_c is very high [for instance, polyvinylalcohol (PVA)-polyacrylic acid hydrogels[30] show $\nu_c \approx 10^{13}$ Hz, close to the maximum frequency allowed in material media]. Therefore, for usual US frequencies the factor $F_{(\nu)}$ in hydrogels should be always used in the low frequency limit (i.e., $F_{(\nu)}=1$). This fact gives the US attenuation[17] proportional to ν^2 while the experimental results follow the law[15] $\propto \nu^n$ with n ranging between 1,25 and 1,50.

In the frame of the bi-phasic approach one possible way out is to consider a different fluid-network interaction for hydrogels.

If we look at the hydrogel structure, it shows water molecules that are bounded to the matrix polymer chains by means of chemical interactions. The local polymer field organizes the water molecules around itself in a way that they can be assumed radially bounded to its chains, while relatively free to move along the perpendicular direction in a viscous manner. Due to the fact that in an isotropic gel the polymer chains are oriented in all directions, under the elastic wave action, the bounded water molecules will respond in an elastic and viscous manner as a mean. Thence, a more appropriate matrix-fluid interaction scheme is assumed:

(i) a viscoelastic interaction between the bounded water and the polymer matrix (with an elastic constant k and a friction coefficient η) and
(ii) a pure viscous interaction between the bounded water molecules (surrounding the polymer network) and the free water.

A. Compressional waves for dilute poroelastic media

Since the percentile content of water in gels is often very high (up to 99%), in the low frequency limit the poroelastic theories for hydrogels make historically use of the dilute matrix approximation[30–33] [i.e., $\beta \approx 1$, where $\beta = V_w/(V_w + V_p)$, where V_w and V_p are the volumes of water and polymer, respectively]. Such an approximation works very well even for gels with a remarkable polymer content (owing β down to 0.50).

For the high frequency limit, the dilute matrix approximation does not work so well especially for the US wave speed that shows to be sensibly dependent on the polymer content for values of β up to 0.7. Nevertheless, the dilute matrix approximation is still physically meaningful for β bigger than 0.8 where it clearly defines the behavior of the US wave speed.

Before introducing the specific hydrogel fluid-matrix interaction, we derive the poroelastic longitudinal wave equations in the limit of very dilute matrix with incompressible solid and liquid constituents.

In this case the compressional elastic moduli of the fluid and the solid matrix can be assumed to be greater than all the other elastic moduli in the poroelastic longitudinal wave equations[16,17,23,24] that read

$$\nabla^2(P\varepsilon_{\alpha\alpha}+Qe_{\alpha\alpha})=\partial^2(\rho_{11}\varepsilon_{\alpha\alpha}+\rho_{12}e_{\alpha\alpha})/\partial t^2 + \beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{1}$$

$$\nabla^2(Q\varepsilon_{\alpha\alpha}+Re_{\alpha\alpha})=\partial^2(\rho_{12}\varepsilon_{\alpha\alpha}+\rho_{22}e_{\alpha\alpha})/\partial t^2 - \beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{2}$$

where ε_{ij} is the solid strain tensor, $e_{\alpha\alpha}$ is the trace of the liquid strain tensor, β is the pore volume fraction (equating the fluid volume fraction), and ρ_{11}, ρ_{12}, and ρ_{22} are the mass density parameters for the hydrogel that are related to the solid and fluid mass densities ρ_s and ρ_f, respectively, by the identities

$$\rho_{11}+\rho_{12}=(1-\beta)\rho_s,$$

$$\rho_{12}+\rho_{22}=\beta\rho_f,$$

$$\rho_{12}=(1-\alpha)\beta\rho_f,$$

where $\alpha>1$ is a molecular shape factor which is independent of the solid or fluid mass densities. For a matrix built up by spherical particles, Berryman[34] showed to be $\alpha=\frac{1}{2}(\beta^{-1}+1)$.

Furthermore, ρ_{11} is the inertial mass density of the solid matrix when it accelerates while the fluid is contemporarily prevented to move, and ρ_{12} is the mass density parameter for the force that the fluid exerts on the solid as the latter is accelerated relative to the former.[23] This happens since the fluid and the solid in the poroelastic media are coupled to each other. If fluid and solid would be independent of each other (i.e., $\rho_{12}=0$) the inertial mass density of solid and fluid in the bi-phasic medium would result just equal to their specific ones (ρ_s and ρ_f) multiplied by their volume fractions $(1-\beta)$ and β, respectively.

Moreover, P, Q, and R are the poroelastic coefficients of the medium that can be measured by means of jacketed and unjacketed experiments.[24] In the jacketed rheological measurements, the solid matrix is subject to compressional tests carried out maintaining the pore fluid at constant pressure, while in the unjacketed ones, the specimen dilatation is measured as a function of the fluid pressure in the pores, leaving the solid matrix free from external forces.[24]

In the limit of high liquid phase content ($\beta\approx1$) and incompressible constituents, the following conditions over the poroelastic coefficients and mass density parameters hold:[16,17,23]

$$R\gg Q\gg P, \tag{3}$$

$$Q/R\cong(1-\beta)/\beta\cong(1-\beta), \tag{4}$$

$$\rho_{12}=-\rho_{11}\ll\rho_{22} \tag{5}$$

By introducing the above approximations into the longitudinal wave equations we obtain

$$\nabla^2 Qe_{\alpha\alpha}=\partial^2\rho_{11}(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t^2+\beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{6}$$

$$\nabla^2(Q\varepsilon_{\alpha\alpha}+Re_{\alpha\alpha})=\partial^2(\rho_{12}\varepsilon_{\alpha\alpha}+\rho_{22}e_{\alpha\alpha})/\partial t^2 - \beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t. \tag{7}$$

Given that the friction coefficient f is very high[25,27,30–33] it follows that the displacement $|\varepsilon_{\alpha\alpha}-e_{\alpha\alpha}|$ of the "slow" wave (liquid and solid in counter-phase) is very small compared to that of the fast wave $|e_{\alpha\alpha}|$ so that it is possible to pose $|\varepsilon_{\alpha\alpha}|\approx|e_{\alpha\alpha}|$ and, hence

$$(\rho_{12}\varepsilon_{\alpha\alpha}+\rho_{22}e_{\alpha\alpha})\cong(\rho_{22}+\rho_{12})e_{\alpha\alpha}=\beta\rho_f e_{\alpha\alpha}, \tag{8}$$

$$(Q\varepsilon_{\alpha\alpha}+Re_{\alpha\alpha})\cong(R/\beta)e_{\alpha\alpha} \tag{9}$$

that introduced in Eq. (7) leads to

$$\nabla^2\frac{R}{\beta}e_{\alpha\alpha}\cong\partial^2(\beta\rho_f e_{\alpha\alpha})/\partial t^2-\beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t. \tag{10}$$

Assuming for dilute matrices Berryman's[34] formula (that calculates ρ_{11} for coil-like polymer molecules as those ones of the PVA hydrogel), ρ_{11} reads

$$\rho_{11}=(1-\beta)\rho_s-\tfrac{1}{2}(1-\beta^{-1})\beta\rho_f$$

from where for $\beta\cong1$ it follows that $\rho_{11}\cong0$. Hence, Eqs. (6) and (10) approximately read as

$$\nabla^2 Qe_{\alpha\alpha}\cong\beta f F_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{11}$$

$$\nabla^2\frac{R}{\beta}e_{\alpha\alpha}\cong\partial^2(\beta\rho_f e_{\alpha\alpha})/\partial t^2. \tag{12}$$

Moreover, by introducing Eq. (12) into Eq. (11), the following relation between the "slow wave" and the fast wave displacements,

$$F(\nu)\frac{\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})}{\partial t}\cong\frac{1}{f}\frac{Q}{R}\frac{\partial^2(\beta\rho_f e_{\alpha\alpha})}{\partial t^2}, \tag{13}$$

is obtained.

Finally, by introducing Eq. (13) into Eq. (10), at first order in $(1-\beta)$, we obtain

$$\nabla^2\frac{R}{\beta}e_{\alpha\alpha}\cong\beta\partial^2(\beta\rho_f e_{\alpha\alpha})/\partial t^2 \tag{14}$$

that, under the hypothesis of β constant, reads

$$\nabla^2 Re_{\alpha\alpha}\cong\beta^3\partial^2(\rho_f e_{\alpha\alpha})/\partial t^2. \tag{15}$$

Equation (15) represents a purely elastic wave of the first type that propagates in a medium with a speed c according to the law

$$c^2=R/(\rho_f\beta^3)\approx c_f^2/\beta^3, \tag{16}$$

where c_f is the wave velocity in the fluid phase. Since β is smaller than but close to 1 in Eq. (16) we can observe that, in dilute poroelastic media, the compressional wave has a velocity that is close but always higher than that one in the pure fluid phase.

Since Eq. (14) is a completely elastic wave equation, to take into account for the US energy dissipation, the highest

order of approximation must be considered. In order to do that, we put Eq. (6) into Eq. (10) and, with the help of Eq. (4), we obtain the wave equation

$$\nabla^2 \frac{R}{\beta} e_{\alpha\alpha} \cong \beta \partial^2(\beta \rho_f e_{\alpha\alpha})/\partial t^2 + (1-\beta) \times (\rho_{11}/\beta f F_{(\nu)}) \partial^3(\rho_f e_{\alpha\alpha})\, \partial t^3. \quad (17)$$

1. Phase velocity and attenuation of "fast" plane wave

Equation (17) for plane waves $e_{aa} \propto e^{-\alpha x} e^{i(kx-\omega t)}$ gives the characteristic equation

$$(k+i\alpha)^2 = \omega^2 \rho_f (\beta^3/R)[1 - i\omega(Q/R)(\rho_{11}/fF_{(\nu)})]. \quad (18)$$

By solving k and α and by using the purely elastic phase velocity $c_0^2 = R/\rho_f \beta^2$, the speed and the attenuation per cycle are obtained, respectively, to be

$$c^2 = c_0^2(1 + \alpha^2/k^2), \quad (19)$$

$$2\pi\alpha/k = -\pi(c/c_0)^2 \omega(1-\beta)(\rho_{11}/\beta f F_{(\omega)}). \quad (20)$$

B. Compressional fast wave in hydrogels

Introducing a new constituent such as the bounded water, we must pay attention to the definition of the fluid volume fraction of the hydrogel.

Since the bounded water phase is not a fluid phase, it must be subtracted from the water volume fraction β that is the total volume fraction of the water.

The effective free water volume fraction β_e can be obtained on the hypothesis that the number of bounded water molecules is proportional to the polymer-water contacts.

Since the probability of a polymer-polymer contact at very low polymer concentration [near $\beta \approx 1$, let us say for $(1-\Delta) \ll \beta < 1$] is practically null, the bounded water volume fraction β_{bw} is proportional to the polymer volume fraction $(1-\beta)$ in that range.

When the polymer concentration is high [let us say for $0 < \beta \ll (1-\Delta)$], where Δ represents the parameter that marks the boundary between the two regimes, due to the high probability of polymer-polymer contacts, the increase in polymer content does not cause a proportional increase in bounded water so that β_{bw} must approach a constant value ϕ as β goes to zero.

If we approximate the approaching of the bounded water volume fraction to this constant value ϕ by means of an exponential law, we can write

$$\beta_{\mathrm{bw}} = \phi(1 - \exp[-(1-\beta)/\Delta]), \quad (21)$$

where $0 < \Delta < 1$ and $0 < \phi < 1$ are empirical parameters to be deduced from the experimental data.

Therefore, the free water volume fraction reads as

$$\beta_e = \beta - \phi(1 - \exp[-(1-\beta)/\Delta]). \quad (22)$$

By introducing the free water volume fraction β_e, the pure elastic acoustic fast wave (15) and its speed, respectively, read as

$$\nabla^2 R e_{\alpha\alpha} \cong \beta_e^3 \partial^2(\rho_f e_{\alpha\alpha})/\partial t^2, \quad (23)$$

$$c_0^2 \cong \frac{c_f^2}{(\beta - \phi(1 - \exp[-(1-\beta)/\Delta]))^3}, \quad (24)$$

where $c_0 = (R/\rho_f \beta_e^3)^{1/2}$ is the pure elastic longitudinal wave velocity, $c_f = (R/\rho_f)^{1/2}$ its velocity in the intermolecular fluid (free water), R is Biot's compressional modulus of the fluid,[16,21] and $e_{\alpha\alpha}$ is the trace of the free water stress tensor.

It is worth noting that for vanishing Δ and for $\beta \leq (1-\varepsilon)$, where $\varepsilon \gg \Delta$ is a positive number (smaller than 1), it holds that $\beta_{\mathrm{bw}} \cong \phi$, $\beta_e \cong (\beta - \phi)$, and Eq. (24) leads to the relation

$$c_{0(\beta)}^2 \cong \frac{c_f^2}{(\beta - \phi)^3}. \quad (25)$$

Therefore, the amount of bounded water ϕ can be evaluated by means of the best-fitted value $\lim_{\beta \to 1} c_{0(\beta)}$ of the experimental data $c_{0(\beta_{(i)})}$ for $\beta_{(i)} \leq (1-\varepsilon)$ following the equation

$$\phi = 1 - \left(\frac{c_f}{\lim\limits_{\beta \to 1} c_{0(\beta)}} \right)^{2/3}. \quad (26)$$

It is noteworthy to note that the presence of the bounded water generates a velocity gap between the US velocity in pure water and that one extrapolated for hydrogels at $\beta = 1$ (a hydrogel made up by 100% of water).

1. Bounded water-polymer network viscoelastic interaction

The bounded water-polymer network viscoelastic interaction can be introduced in the poroelastic equations by adding to Biot's viscous force (with $F_{(\nu)} = 1$)

$$\beta_e f\, \partial(e^*_{\alpha\alpha} - e_{\alpha\alpha})/\partial t, \quad (27)$$

the viscoelastic one due to the bounded water

$$\eta_{(\nu)}\, \partial(\varepsilon_{\alpha\alpha} - e^*_{\alpha\alpha})/\partial t + \kappa(\varepsilon_{\alpha\alpha} - e^*_{\alpha\alpha}), \quad (28)$$

where $e^*_{\alpha\alpha}$ is the trace of the bounded water stress tensor.

Actually, introducing the bounded water stress tensor variable, the bi-phasic approach disembogues into a three-phasic one that is out of the purpose of this work. Nevertheless, the bi-phasic model can be retained since the bounded water mass density is not much different from that one of the polymer. It is much like the bounded water "inflates" the polymer chains, creating a bearing around them. In this case, the bi-phasic medium can be conceived composed of the "polymer-bounded water aggregate," as matrix, plus the interstitial free water. Thence it is possible to introduce the following approximations.

(i) The mass density of the solid aggregate approximates that one of the polymer.
(ii) The trace of the strain tensor of the polymer $\varepsilon_{\alpha\alpha}$ approximates that one of the solid aggregate.
(iii) P, Q, and R are the poroelastic coefficients of the new bi-phasic medium. Moreover, since not the whole bounded water is involved in the fluid-solid shear process.

(iv) The inertial effect of bounded water involved in the fluid-solid shear process can be disregarded.

By means of the fourth hypothesis, the force between the free water and the bounded water can be equated to that one between the bounded water and the polymer matrix, leading to the additional equation needed to solve the three stresses that reads

$$\beta_e f F_{(\nu)}\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t=\eta_{(\nu)}\,\partial(\varepsilon_{\alpha\alpha}-e^*_{\alpha\alpha})/\partial t+\kappa(\varepsilon_{\alpha\alpha}-e^*_{\alpha\alpha}). \tag{29}$$

Moreover, under hypotheses (i)–(iii) the system of poroelastic equations (1) and (2) reads as

$$\nabla^2(P\varepsilon_{\alpha\alpha}+Qe_{\alpha\alpha})=\partial^2(\rho_{11}\varepsilon_{\alpha\alpha}+\rho_{12}e_{\alpha\alpha})/\partial t^2+\beta f\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{30}$$

$$\nabla^2(Q\varepsilon_{\alpha\alpha}+Re_{\alpha\alpha})=\partial^2(\rho_{12}\varepsilon_{\alpha\alpha}+\rho_{22}e_{\alpha\alpha})/\partial t^2-\beta f\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t. \tag{31}$$

In dilute bi-phasic media the compressional plane wave of interstitial fluid, $e_{\alpha\alpha}\propto e^{-\alpha x}e^{i(kx-\omega t)}$, induces a planar slow wave, as shown in Eq. (13), that leads to

$$(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})=-(i\omega Q\beta_e{}^2/Rf)e_{\alpha\alpha},$$

where $\omega=2\pi\nu$.

In this case Eq. (29) reads as

$$\beta_e f\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t=(\eta_{(\nu)}+j(\kappa/\omega))\,\partial(\varepsilon_{\alpha\alpha}-e^*_{\alpha\alpha})/\partial t, \tag{32}$$

leading after simple manipulation to the relation

$$\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t=\{(\beta_e f+\eta_{(\nu)}+j(\kappa/\omega))/(\eta_{(\nu)}+j(\kappa/\omega))\}\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{33}$$

where that can be recast as

$$\beta_e f\,\partial(e^*_{\alpha\alpha}-e_{\alpha\alpha})/\partial t=F^*_{(\omega)}\,\partial(\varepsilon_{\alpha\alpha}-e_{\alpha\alpha})/\partial t, \tag{34}$$

where the complex "friction" coefficient $F^*_{(\omega)}$ reads as

$$F^*_{(\omega)}=[(\beta_e f)^{-1}+(\eta_{(\nu)}+j(\kappa/\omega))^{-1}]^{-1}. \tag{35}$$

Hence, by introducing Eq. (34) into Eqs. (30) and (31) for plane waves, the dilute matrix equation (17) reads as

$$\nabla^2\frac{R}{\beta}e_{\alpha\alpha}\cong\beta\partial^2(\beta\rho_f e_{\alpha\alpha})/\partial t^2+(1-\beta)\times(\rho_{11}/F^*_{(\omega)})\partial^3(\rho_f e_{\alpha\alpha})/\partial t^3 \tag{36}$$

so that for the fast plane wave $e_{\alpha\alpha}\propto e^{-\alpha x}e^{i(kx-\omega t)}$ it leads to the characteristic equation

$$(k+i\alpha)^2=\omega^2\rho_f(\beta_e{}^3/R)[1-i\omega(1-\beta_e)(\rho_{11}/F^*_{(\omega)})] \tag{37}$$

that solved in α and c gives

$$c^2=c_0^2\left(1+\alpha^2/k^2\right)\Big/\left[1+(1-\beta_e)\omega\rho_{11}\,\mathrm{Im}\{F^{*-1}_{(\omega)}\}/\beta_e{}^2\right], \tag{38}$$

$$\alpha/k=-1/2(c/c_0)^2\omega(1-\beta_e)\rho_{11}\,\mathrm{Re}\{F^{*-1}_{(\omega)}\}, \tag{39}$$

where

$$\mathrm{Re}\{F^{*-1}_{(\omega)}\}=[\eta_{(\omega)}(1+\eta_{(\nu)}/\beta_e f)+(\kappa^2/\beta_e f\omega^2)]/(\eta_{(\nu)}{}^2+(\kappa/\omega)^2), \tag{40}$$

$$\mathrm{Im}\{F^{*-1}_{(\omega)}\}=-(\kappa/\omega)/(\eta_{(\nu)}{}^2+(\kappa/\omega)^2). \tag{41}$$

In the following we consider the case when the bounded water-network interaction is prevalently viscous such as

$$\eta>(\kappa\rho_{\mathrm{bw}})^{1/2}, \tag{42}$$

where ρ_{bw} is the mass density of the bounded water that we assume close to that of free water ρ_f.

At high frequencies such $\omega>\omega_g=(2\pi\eta/\rho_f)>(\kappa/\rho_{\mathrm{bw}})^{1/2}$ (far away from bounded water resonance) hence, it holds that

$$\kappa/\omega\eta\ll\kappa/\omega_g\eta<1. \tag{43}$$

Moreover, in order that the crossover frequency $\nu_g=\eta/\rho_f$ is much smaller than the unattainable Biot's one $\nu_c=f/\rho_{f'}$ in the high frequency limit, it must hold

$$\lim_{\nu\to\infty}\eta_{(\nu)}\ll f. \tag{44}$$

If the high frequency behavior of the bounded water viscosity $\eta_{(\nu)}$ is modeled by a dimensionless coefficient as

$$\eta_{(\nu)}=\eta_0(\omega_g/\omega)^{\delta}, \tag{45}$$

it follows that $\omega_g=2\pi\nu_g=2\pi\eta_{(\nu=\nu_g)}/\rho_f=2\pi\eta_0/\rho_{f'}$ and that conditions (43) and (44) hold contemporarily by requiring $0<\delta\leq 1$. In such a case, by introducing them into Eqs. (40) and (41) it follows that

$$\lim_{\omega/\omega_g\gg 1}\mathrm{Re}\{F^{*-1}_{(\omega)}\}\cong(\kappa/\omega\eta)^2+[(1+\eta_{(\nu)}/\beta_e f)/\eta_{(\nu)}]\cong 1/\eta_{(\nu)}, \tag{46}$$

$$\lim_{\omega/\omega_g\gg 1}\mathrm{Im}\{F^{*-1}_{(\omega)}\}\cong 0, \tag{47}$$

where $\nu=2\pi/\omega$. Equations (46) and (47) introduced into Eq. (38), in a dilute polymer hydrogel ($\beta\approx 1$ and $\rho_{11}\ll 1$), lead to

$$c^2=c_0^2(1+\alpha^2/k^2)\cong c_0^2, \tag{48}$$

being typically $(\alpha/\kappa)^2$ very small (of order of 10^{-3} in our experiments) and to the specific US attenuation per cycle

$$2\pi\alpha/k\approx-\pi(c/c_0)^2(\omega/\omega_g)^{1+\delta}(1-\beta+\phi(1-\exp[-(1-\beta)/\Delta])\sigma_{pf}, \tag{49}$$

where $\sigma_{pf}=\rho_{11}/\rho_f$.

For hydrogels with small Δ, and for $(1-\beta)/\Delta\gg 1$, finally, Eq. (49) leads to

$$2\pi\alpha/k\approx-\pi(c/c_0)^2(\omega/\omega_g)^{1+\delta}(1-\beta+\phi)\sigma_{pf}. \tag{50}$$

From Fig. 1, that shows the acoustic US fast wave absorption, it can be observed that the bounded water raises the linear behavior, while Δ smooths the curve in the interval $1-\Delta>\beta>1$.

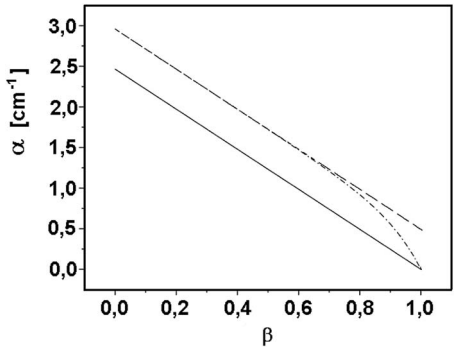


FIG. 1. Theoretical behavior of acoustic US fast wave absorption in a hydrogel given by Eq. (52) as a function of the water volume fraction for various values of the parameters: $\phi=0$ and $\Delta=0$ (full line), $\phi=0.2$ and $\Delta=0.0006$ (dashed line), and $\phi=0.2$ and $\Delta=0.1$ (dashed-dotted line).

From Fig. 2, depicting the theoretical behavior of the acoustic speed of US fast wave, it can be seen that

(i) the bounded water leads to an upward shift in the wave speed,
(ii) that Δ smooths the curve upward in the region $1-\Delta<\beta<1$, and
(iii) when $\Delta \ll 1$ there is a speed jump at $\beta \cong 1$.

C. High polymer concentration

The characteristic equation for slow and fast plane waves stemming from Eqs. (30) and (31) turns out to be formally the same as that one given by Biot[17] where the friction term $\beta f F_{(\nu)}$ is substituted by $F^*_{(\omega)}$ [see Eq. (34)] that at high frequency is given by Eqs. (46) and (47).

Nevertheless, it must be observed that Biot's[17] model assumes that the poroelastic coefficients are constants (or at least very smoothly varying) while in hydrogels they actually may change very much when they are partially dried up.[33]

Moreover, since we wish both to build up a model that can be applied to biological tissues, where the solid concentration as well as that one of the bounded water can be very high (up to 60%), and to improve the experimental fitting of data beyond the range of validity of the dilute matrix approximation at intermediate water volume fraction values (but not to validate the model), we opt for a more direct semi-empirical approach.

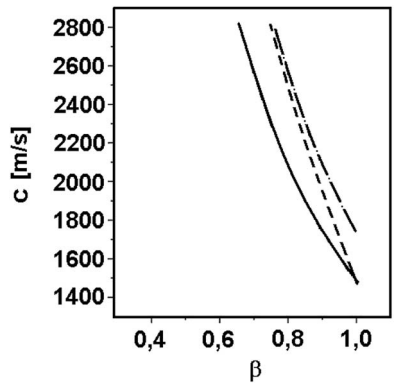


FIG. 2. Theoretical behavior of the acoustic speed of propagation of US fast wave in a hydrogel given by Eq. (24) as a function of the water volume fraction for various values of the parameters: $\phi=0$ and $\Delta=0$ (full line), $\phi=0.1$ and $\Delta=0$ (dashed line), and $\phi=0.1$ and $\Delta=0.1$ (dashed-dotted line).

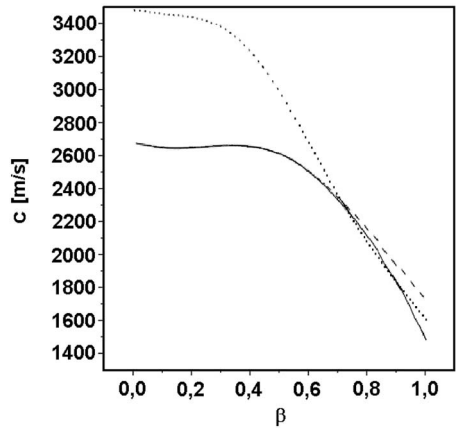


FIG. 3. Theoretical behavior of the acoustic speed of propagation of US fast wave in a hydrogel given by Eq. (52) as a function of the water volume fraction for various values of the parameters: $\phi=0.1$, $\Delta=0$, $\gamma=0.7$, and $\chi=-0.4$ (dashed line); $\phi=0.1$, $\Delta=0.1$, $\gamma=0.7$, and $\chi=-0.4$ (full line); $\phi=0.031$, $\Delta=0$, $\gamma=0.44$, and $\chi=-0.25$ (dotted line).

In order to do that, we observe that for vanishing $\beta=0$ values, the speed of propagation of the US is finite and equal to that of polymeric solid c_s, while the law (24) diverges to infinity at $\beta=\phi$ and gives negative values for $\beta=0$.

The contribution that cancels the divergence of the denominator of Eq. (24) (let us call it $G_{(\beta)}$ in the dilute matrix approximation) can be put in the form

$$c^2 = \frac{c_f^2}{(\beta - \phi(1-\exp[-(1-\beta)/\Delta]))^3 + G_{(\beta)}}, \tag{51}$$

where the conditions required to $G_{(\beta)}$ are

$$\lim_{\beta\to 1} G_{(\beta)} = 0,$$

$$\lim_{\beta\to 0} G_{(\beta)} > 0.$$

For this reason it is appropriate to use the series approximation

$$G(\beta) = \gamma(1-\beta) + \chi(1-\beta)^2 \quad \text{with} \quad \gamma+\chi>0$$

that in Eq. (29) leads to

$$c^2 = \frac{c_f^2}{(\beta - \phi(1-\exp[-(1-\beta)/\Delta]))^3 + \gamma(1-\beta) + \chi(1-\beta)^2} \tag{52}$$

and that for $(1-\beta)/\Delta \gg 1$ gives

$$c^2 = \frac{c_f^2}{(\beta-\phi)^3 + \gamma(1-\beta) + \chi(1-\beta)^2}. \tag{53}$$

From Fig. 3 we can see the change in the US wave speed following the introduction of the polynomial expression $G_{(\beta)}$: the polymer network cancels the divergence of US speed at $(\beta-\phi)=0$ and lowers it at the intermediate values of β in a progressive manner.

Finally, for β close to 1 we observe that the increase in the bounded water in Eq. (53) increments the US speed through the term $(\beta-\phi)^3$ while the elasticity of the polymer network lowers it mainly through the vanishing term $\gamma(1-\beta)$ (if $\gamma>0$).

III. EXPERIMENTAL

A. Materials and methods

Gel samples were prepared in parallelepipeds of $2\times2\times1$ cm^3 using PVA with a degree of hydrolysis of 99+% and an average molecular weight of 115.000 ± 30.000 (Aldrich, Milan, Italy) dissolved into de-ionized water in a concentration of 10% by weight. The homogeneous solution was refrigerated at −80 °C starting from room temperature 23 ± 1 °C. The freezing-thawing procedure was repeated two, three, four, and six times. The samples were left to equilibrate in de-ionized water for 72 h.

The experimental tests were carried out at different hydration conditions down to a minimum of about 50%, taking care that the drying was gradual and homogeneous.

The reversibility of the de-hydration treatment was checked at the end after the drying steps.

The ultrasonic pulses were generated by a Panametrics® Pulser model 5052PR (Waltham, MA) coupled with a polyvinylidene fluoride piezoelectric transducer assembled in our laboratory following the procedure of Naganishi *et al.*[35]

A cylindrical chamber immersed in a water bath constitutes the experimental cell. The US transducer/receiver is fixed on a planar side-wall. An inner flat counter-wall, free to translate inside the cylinder and put in contact with the back surface of the samples, is used as reflecting surface.

The ultrasonic disk-like transducer used in the experiment has an aperture size of 1.5 cm. The measurements are carried out in the near field condition with the front surface of the gel samples in contact with the transducer surface. All the volume in front of the transducer is occupied by the sample (so that there is no problem about US focalization) and US echoes are very well detected up to 10 cm far apart from the transducer.

The Pulser voltage spike of −270 V induces an ultrasonic pulse of 250 kHz of frequency through the transducer.

The distance between the transducer and the reflecting stainless steel layer behind the samples was measured with an accuracy of ±0.01 cm and maintained constant through the whole experiment.

Echo signal registration and conditioning data were collected with a routine and carried out with the LABVIEW™ software on a computer through a National Instruments® data acquisition device.

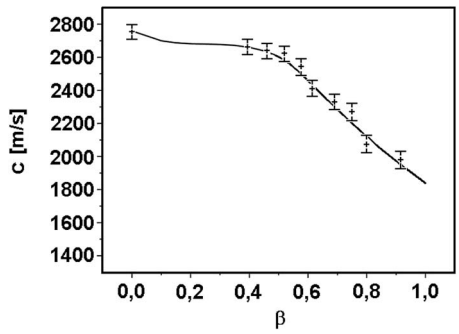


FIG. 4. Experimental acoustic US speed of propagation of fast wave in PVA hydrogels, with best fit (full line), as a function of the water volume fraction $(0<\beta<1)$ for a hydrogel sample obtained by means of 2 cycles of cross-linking.

Finally, the samples were totally dehydrated in an oven at 40 °C with desiccant silica gels to measure their polymer content and the US speed in the dry solid.

The US absorption coefficient α was deduced by using the mathematical relation $\alpha=(1/2d)\ln(A_0/A_{(2d)})$, where A_0 and $A_{(2d)}$ represent both the initial and final wave amplitudes, respectively, and where d is the sample thickness.

The water volume fraction of the hydrogel samples β was obtained by means of the respective weight fractions P_w and P_p such as $\beta\approx P_w/(P_w+P_p)$ since the water and PVA specific densities are very close to each other.

After the samples were left to fully hydrate themselves in distilled water, the fractional water volume was measured to be $\beta=0.917\pm0.001$ for all the samples.

The fitting of the experimental results was carried out by means of a multiple parameter best fit utilizing an appropriate routine in MATLAB® 7.0.

The error bars are obtained both from theoretical calculation (taking into account all the experimental errors in space, time, and temperature) and from experimental evaluation. The mean square root deviation of data resulted much smaller than the theoretical errors.

B. Ultrasound phase velocity

Experimental data of US wave speed with the best-fitted curve are shown in Fig. 4 for the hydrogel sample submitted to 2 cycles of cross-linking. The most probable values of the model parameters for all samples are given in Table I.

TABLE I. Most probable values of the model parameters ϕ, Δ, γ, χ, and $\gamma+\chi$ obtained from the best fit of the experimental data by means of expression (31) as a function of the number of cross-linking cycles (first column).

Number of cross-linking cycles	ϕ	Δ	γ	χ	$\gamma+\chi$	$\lim_{\beta\to1} c$ (m s^{-1})	c_s (m s^{-1})
2	0.131	0.0006	0.742	−0.453	0.289	1830	2755
3	0.075	0.0006	0.570	−0.363	0.206	1667	3266
4	0.063	0.0006	0.445	−0.247	0.197	1636	3339
6	0.020	0.0006	0.307	−0.109	0.198	1528	3340

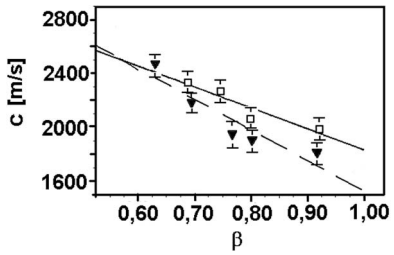


FIG. 5. Experimental acoustic US speed of propagation of fast wave in PVA hydrogels as a function of the water volume fraction at various degrees of polymer matrix cross-linking: 2 cycles (⊡) and 6 cycles (▽) of cross-linking, with the best fit (dashed line and full line, respectively). In this case, the detail of the US speed behavior for high water volume fraction values ($0.6<\beta<1$) is shown.

From Fig. 4 we can see that at intermediate values of β, the US wave speed progressively departs from the model dependence $(\beta-\phi)^{-3/2}$ toward the value of the dry solid polymer in agreement with the theoretical behavior (52) shown in Fig. 3. The measurements show that the effect of the polymer matrix starts to influence the US speed of propagation for values of β as high as 0.8.

For β close to zero, the US speed of propagation converges to the speed of the dry polymer which rises when the number of cross-linking cycles increases because of the greater polymer stiffness, as shown in Table I.

For β equal to 1, the US speed is extrapolated by means of a best-fit procedure and reported in Table I. The data referring to the samples with 2 and 6 cycles of cross-linking have been reported in Fig. 5. The data show that the US phase velocity is sensibly higher than that of pure water (that in our experimental condition has been measured to be 1483 m/s). In Fig. 6 the bounded water volume fraction ϕ is shown as a function of the number of cross-linking cycles of the sample. The evaluation of ϕ is obtained by introducing into Eq. (26), being $\Delta\cong 0$, the limiting velocity values of Table I. The bounded water volume fraction ϕ ranges from 2% in the PVA samples with six cross-linking cycles to 13% for the PVA samples with two cross-linking cycles.

As shown in Table I, we can observe that the lower the gap speed at $\beta=1$, the bigger the number of the cross-linking cycles and the lower the measured bounded water fraction ϕ present in the hydrogel. This agrees with the characteristics of PVA hydrogel synthesized by means of freezing-thawing

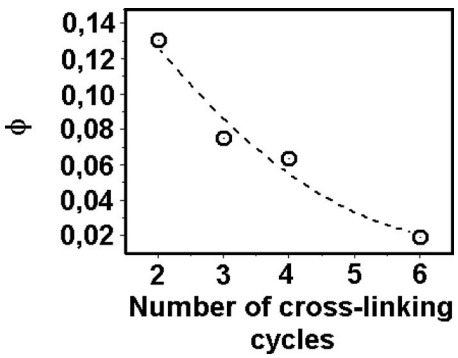


FIG. 6. Estimated bounded water fraction in PVA hydrogel as a function of the number of cross-linking cycles.

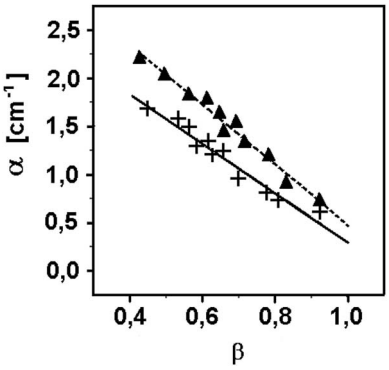


FIG. 7. Experimental acoustic US fast wave attenuation in PVA hydrogels as a function of the water volume fraction at various degrees of polymer matrix cross-linking: 4 cycles (▲) and 6 cycles (+) of cross-linking with best fit (dashed-dotted line and dotted line, respectively).

cycles where the number of polymer-polymer contacts (cross-links) grows as the number of freezing-thawing cycles increases.[36] Moreover, since lower values of ϕ in samples with higher cross-linking correspond to lower values of the US speed, this behavior cannot be ascribed to the elasticity of the polymer network that leads to a variation in opposite sign.

In fact, as shown in Table I, both γ and $(\gamma+\chi)$ (that account for the effect of the polymer network elasticity on the US speed [γ contributes prevailing in the region of high $\beta\approx 1$, while $(\gamma+\chi)$ in the region of small $\beta\approx 0$]) decrease as the network cross-linking increases so that by looking at Eq. (53) it comes clearly out that in such a case they would cause an increase in the US speed.

Moreover, since both γ and $(\gamma+\chi)$ are positive for the intermediate values of β, the effect of the polymer network is to decrease the US speed and to eliminate the US speed divergence due to the dilute matrix approximation for small values of β.

From experimental data we can also see that the value of the parameter Δ obtained from the last square fit procedure is equal to 0.0006 (practically null) for all PVA gel samples so that ϕ effectively represents the bounded water concentration and $(\beta-\phi)$ the free water concentration. This fact says that the bounded water concentration remains practically unchanged inside the PVA gel samples (as a function of β) so that primarily free water is subtracted when they are (mildly) dried.

C. Ultrasound wave attenuation

Some of the experimental data as well as the fits of the theoretical expression (49) are shown in Fig. 7. For the attenuation measurements as well as for the velocity ones, the value of the fitted parameter Δ resulted practically null.

A linear decrease in the US wave attenuation α is measured as a function of β for all samples with a mean value $\partial\alpha/\partial\beta=-2.4\pm 0.4$ cm^{-1}.

The pure water attenuation has also been measured to be 0.022 cm^{-1} and found to be much smaller than that one of the gel in the order of 1 cm^{-1}.

IV. DISCUSSION

The US propagation in natural hydrogels and soft tissues is currently described by a single-phase wave equation[37,38] where the medium is conceived as a fluid "solution" in which dissolved molecules can introduce processes of both dissipation and resonance. Even if this scheme completely ignores the liquid-solid arrangement, it gives appreciable results that are able to explain many aspects of tissue acoustical behavior. This probably happens because of the high built-in flexibility of such an equation that is able to comprehend a wide class of acoustical media thanks to a number of parameters that can be empirically adjusted to describe the propagation of acoustic waves.

Yang and Cleveland[39] showed that it is possible to reproduce the frequency dependence of the absorption with an exponent equal to 1.1 by supposing two juxtaposed relaxing processes that may or may not be real. In this way some information about the material medium (e.g., a natural gel or a tissue) is lost from the US behavior.

The model presented here discloses how the behavior of the US propagation is linked to the arrangement of the solid and liquid phases of the medium.

More recently Kowalski[40] developed a theory for the propagation of US in a dilute suspension. Even if conceptually different from a polymeric gel, a fluid suspension can be assimilated to a bi-phasic medium with a solid matrix (with its own elasticity and permeability) and a fluid that permeates the interstices. Obviously, the Kowalski model is expressed in terms of the viscoelastic parameters of the suspension that do not correspond *tout court* to those of the bi-phasic approach, except for the fluid fraction volume β.

The theoretical dependence of the US velocity as a function of the water volume fraction found by Kowalski is of the same type obtained in the present paper with the same good experimental agreement. The main difference is that the suspension does not have bounded water (i.e., $\phi=0$) so that at $\beta=1$ the US speed converges to that of the pure fluid.

Hence, the following features of the US propagation in hydrogels can be ascribed to their state of aggregation.

(i) The speed of propagation of US in highly hydrated gels is always a little bit higher than that of water (1480 m/s) since in Eq. (26) $\beta<1$ and $\phi>0$.

(ii) The US specific attenuation $2\pi\alpha/\kappa$ in natural hydrogels can follow a non-integer frequency law $\nu^{(1+\delta)}$ with $0<\delta\leq 1$, where δ is the exponent of the polymer-bounded water viscous interaction.

This outcome well agrees with the attenuation data of US in natural gels where δ is typically about 0.25–0.50 while the value for pure water is 1.

As far as it concerns the applicability of the diluted matrix approximation, we observe that the polymer content sensibly influences the US wave speed up to values of β as high as 0.7. For β higher than 0.70 the polymer effect on the US speed becomes smaller and smaller and the behavior approaches that one of dilute gels depicted in Fig. 1. By using Eq. (53) for $\beta=0.917$, the variation in the US speed due to the polymer network can be evaluated to be about 3.5% (for the PVA samples submitted to four cross-linking cycles) with a maximum value of 5.2% for the samples submitted to two cross-linking cycles.

It must also be noted that the development of tissue-mimicking US phantoms can take advantage by a model that discloses the relation between the medium organization and the US behavior.

In principle, even though living tissues, or organs, have their own complex morphology and anisotropy, a simplified bi-phasic model of soft (isotropic) natural tissue can be conceived by modeling biological cells as poroelastic spheres, endowed by internal and superficial elasticity and permeability, dispersed in a hydrogel environment.

Although the polymer and the bounded water contents of living tissues can reach much higher concentrations than in PVA gels, the present model can be used to describe the behavior of each type of biological hydrogel of a soft tissue (e.g., extra-cellular matrix, internal body of the cell, and so on). In this case, the "free water fractional content" is given by the effective β [given by Eq. (22) with $\Delta=0$] while the polymer-bounded water aggregate constitutes the solid part of the hydrogel matrix.

Even considering this simplest case, there are some limitations derived from the increased complexity of the architecture of the medium. First of all, in order to maintain the continuum approach, the US wave must not resolve the structure of the single cell. As a consequence of this, the theory is limited to US frequencies whose wavelength is much bigger than the cell dimensions. For instance, for cells of order of 30 μm and wave speed of order of 2×10^3 m/s, the upper frequency limit for the theory is smaller than 66 MHz.

It must also be pointed out that the number of material parameters of the model increases. In addition to β and f of the extra-cellular medium, the model would require β and f of the inner volume of the cells, the superficial elasticity as well as permeability of their membrane, and the percentage of the cellular volume. Such a relatively large number of material parameters will give rise to many possible tissue configurations. In this case, the major problem is to identify the appropriate set of values of those coefficients for each real tissue and how experimentally to measure them. Although the complexity of the model increases, it can definitely describe the US propagation in terms of collective cells and extra-cellular matrix poroelastic characteristics.

This opens up the way to US non-invasive tissue physiology evaluation since the permeability and elasticity of the cell structures depend on their state and functionality.

Once the model is validated, then it could be refined by introducing capillaries and blood vessels in the tissue as tubes filled by fluid (with their own internal β and wall elasticity and permeability).

V. CONCLUSIONS

In the present work a continuum model for US acoustic longitudinal waves in hydrogels has been presented. The model shows that a speed gap, at the hydrogel water volume fraction $\beta=1$, is generated by the presence of the bounded

water around the hydrogel polymer matrix: the bigger the bounded water concentration, the higher the US speed gap at $\beta=1$.

The poroelastic model also shows that in natural hydrogels the US attenuation follows a non-integer frequency law $\nu^{(1+\delta)}$ with $0<\delta\leq 1$ depending on the exponent δ describing the polymer-bounded water viscous interaction.

The experimental measurements confirm the gap between the US speed of propagation in the gel samples and that one of pure water at β close to 1 (100% water). Furthermore, the gel samples with a lower degree of cross-linking (and hence, with higher bounded water volume fraction) show a higher US speed gap at $\beta=1$. The evaluated bounded water fraction ranges from 2% in the PVA samples with six cross-linking cycles to 13% for the PVA samples with two cross-linking cycles.

In the range of experimental conditions ($\beta>0.4$), the bounded water volume fraction ϕ inside each type of PVA hydrogel does not depend on the free water amount. Moreover, as the gel water fraction β is lowered toward zero, the US propagation speed is more and more influenced by the matrix elasticity and increases in samples with higher network cross-linking in agreement with the classical rheological law.

The experimental data also show that the US attenuation in hydrogels decreases with the increase in the water volume fraction β in a linear way in agreement with the theoretical forecast.

All these results have been obtained by introducing into the bi-phasic theory the interaction of the bounded water around the polymer network that is peculiar of hydrogels. As a consequence of that, the proposed bi-phasic model has the capability to describe more appropriately the US propagation both in artificial hydrogels and in soft biological tissues, resulting in a promising tool in the biomedical field either for the development of non-invasive tissue-typing techniques or for the realization of phantoms for experimental tests such as those regarding the thermal therapy.

[1] J. C. Bacri and R. Rajaonarison, "Ultrasonic study of the critical phenomena in gels," J. Phys. (France) Lett. **40**, 15–18 (1979).

[2] J. P. Jarry and G. D. Patterson, "Hypersonic relaxation in polybutadiene and polyisoprene," J. Polym. Sci., Polym. Phys. Ed. **19**, 1791–1797 (1981).

[3] J. C. Bacri, J. M. Courdille, J. Dumas, and R. Rajaonarison, "Ultrasonic waves: A tool for gelation process measurements," J. Phys. (France) Lett. **41**, 369–372 (1980).

[4] K. Yuan, Z. Hu, and Y. Li, "Polymer gel as thermally responsive attenuator for ultrasonic waves," Appl. Phys. Lett. **74**, 2233–2235 (1999).

[5] M. L. Mather and C. Baldock, "Ultrasound tomography imaging of radiation dose distributions in polymer gel dosimeters: Preliminary study," Med. Phys. **30**, 2140–2148 (2003).

[6] F. A. Duck, *Acoustic Properties of Tissue at Ultrasonic Frequencies* (Academic, London, 1990), Chap. 4, pp. 75–99.

[7] F. W. Kremkau, R. W. Barnes, and C. P. McGraw, "Ultrasonic attenuation and propagation speed in normal human brain," J. Acoust. Soc. Am. **70**, 29–38 (1981).

[8] J. W. Wladimiroff, I. L. Craft, and D. G. Talbert, "In vitro measurements of sound velocity in human fetal brain tissue," Ultrasound Med. Biol. **1**, 377–382 (1975).

[9] D. De Rossi, A. Nannini, and C. Domenici, "Artificial sensing skin mimicking mechanoelectrical conversion properties of human dermis," IEEE Trans. Biomed. Eng. **35**, 83–92 (1988).

[10] S. Lochhead, D. Bradwell, R. Chopra, and M. J. Bronskill, "A gel phantom for the calibration of MR-guided ultrasound thermal therapy," in 2004 IEEE Ultrasonics Symposium, Montreal, Canada (2004), Vol. **2**, pp. 1481–1483.

[11] G. Divkovic, M. Liebler, K. Braun, T. Dreyer, P. Huber, and J. Jenne, "Thermal properties and changes of acoustic parameters in an egg white phantom during heating and coagulation by high intensity focused ultrasound," Ultrasound Med. Biol. **33**, 981–986 (2007).

[12] M. Ziol, A. Handra-Luca, A. Kettaneh, C. Christidis, F. Mal, F. Kazemi, V. de Lédinghen, P. Marcellin, D. Dhumeaux, J.-C. Trinchet, and M. Beaugrand, "Noninvasive assessment of liver fibrosis by measurement of stiffness in patient with chronic hepatitis C," Hepatology (Philadelphia, PA, U. S.) **41**, 48–54 (2005).

[13] G. P. Berry, J. C. Bamber, C. G. Armstrong, N. R. Miller, and P. E. Barbonne, "Toward an acoustic model-based poroelasticity imaging method: I. Theoretical foundation," Ultrasound Med. Biol. **32**, 547–567 (2006).

[14] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 396–409 (2004).

[15] F. A. Duck, *Acoustic Properties of Tissue at Ultrasonic Frequencies* (Academic, London, 1990), Chap. 4, pp. 112–113.

[16] M. A. Biot, "General theory of three-dimensional consolidation," J. Appl. Phys. **12**, 155–164 (1941).

[17] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. High frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).

[18] D. L. Johnson, D. L. Hemmick, and H. Kojima, "Probing porous media with first and second sound. I. Dynamic permeability," J. Appl. Phys. **76**, 104–114 (1994); D. L. Johnson, T. J. Plona, and H. Kojima, "Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media," *ibid.* **76**, 115–125 (1994).

[19] D. L. Johnson and S. Kostek, "A limitation of the Biot–Gardner theory of extensional waves in fluid-saturated porous cylinders," J. Acoust. Soc. Am. **97**, 741–744 (1995).

[20] E. R. Hughes, T. G. Leighton, G. W. Petley, P. R. White, and R. C. Chivers, "Estimation of critical and viscous frequencies for Biot theory in cancellous bone," Ultrasonics **41**, 365–368 (2003).

[21] D. L. Johnson, "Theory of frequency dependent acoustics in patchy-saturated porous media," J. Acoust. Soc. Am. **110**, 682–694 (2001).

[22] J. Toms, T. M. Muller, R. Ciz, and B. Gurevich, "Comparative review of theoretical models for elastic wave attenuation and dispersion in partially saturated rocks generation," Soil Dyn. Earthquake Eng. **26**, 548–565 (2006).

[23] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).

[24] M. A. Biot, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).

[25] D. L. Johnson, "Elastodynamics of gels," J. Chem. Phys. **77**, 1531–1539 (1982).

[26] R. N. Chandler, "Transient streaming potential measurements on fluid-saturated porous structures: An experimental verification of Biot's slow wave in the quasi-static limit," J. Acoust. Soc. Am. **70**, 116–121 (1981).

[27] A. Peters and S. J. Candau, "Kinetics of swelling of spherical and cylindrical gels," Macromolecules **21**, 2278–2282 (1988).

[28] D. L. Johnson, "Equivalence between fourth sound in liquid He II at low temperature and the Biot slow wave in consolidated porous media," Appl. Phys. Lett. **37**, 1065–1067 (1980).

[29] R. C. Lee, E. H. Frank, A. J. Grodzinsky, and D. K. Royalance, "Oscillatory compressional behavior of articular cartilage and its associated electromechanical properties," J. Biomech. Eng. - Trans. ASME **103**, 280–292 (1981).

[30] P. Chiarelli and D. De Rossi, "Determination of mechanical parameters related to the kinetics of swelling in an electrically activated contractile gel," Prog. Colloid Polym. Sci. **78**, 4–8 (1988).

[31] T. Tanaka and D. J. Fillmore, "Kinetics of swelling of gels," J. Chem. Phys. **70**, 1214–1218 (1979).

[32] P. Chiarelli and D. De Rossi, "Modeling and mechanical characterization of thin fibers of contractile polymer hydrogel," J. Intell. Mater. Syst. Struct. **3**, 396–417 (1992).

[33] P. Chiarelli, C. Domenici, and G. Genuini, "Crazing dynamics in the swelling of thermally cross-linked PVA-PAA films," J. Mater. Sci.: Mater. Med. **4**, 5–11 (1993).

[34]J. G. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **37**, 382–384 (1980).
[35]T. Naganishi, M. Suzuki, and H. Ohigashi, "Ultrasonic transducers," U.S. Patent No. 4,296,349 (20 October 1981).
[36]M. Suzuki and O. Hirasa, "An approach to artificial muscle using polymer gels formed by micro-phase separation," *Advances in Polymer Science* (Springer-Verlag, Berlin, 1993), Vol. **110**. pp. 241–261.
[37]E. A. Zabolotskaya and R. V. Khokhlov, "Quasi-plane waves in nonlinear acoustic of confined beams," Sov. Phys. Acoust. **15**, 35–40 (1969).
[38]V. P. Kuznetsov, "Equation of nonlinear acoustic," Sov. Phys. Acoust. **16**, 467–470 (1970).
[39]X. Yang and R. O. Cleveland, "Time domain simulation of nonlinear acoustic beams generated by rectangular piston with application to harmonic imaging," J. Acoust. Soc. Am. **117**, 113–123 (2005).
[40]S. J. Kowalski, "Ultrasonic waves in diluted and densified suspensions," Ultrasonics **43**, 101–111 (2004).

Self-demodulation of high-frequency ultrasound

Hendrik J. Vos, David E. Goertz,[a] and Nico de Jong[b]
Biomedical Engineering, Thorax Centre, Erasmus MC, Room Ee2302, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

(Received 29 December 2008; revised 16 December 2009; accepted 1 January 2010)

High-frequency (>10 MHz) ultrasound is used in, e.g., small animal imaging or intravascular applications. Currently available ultrasound contrast agents (UCAs) have a suboptimal response for high frequencies. This study therefore investigates the nonlinear propagation effects in a high-frequency ultrasound field (25 MHz) and its use for standard UCA and diagnostic frequencies (1–3 MHz). Nonlinear mixing of two high-frequency carrier waves produces a low-frequency wave, known as the self-demodulation or parametric array effect. Hydrophone experiments showed that the self-demodulated field of a focused 25 MHz transducer (850 kPa source pressure) has an amplitude of 45 kPa at 1.5 MHz in water. Such pressure level is sufficient for UCA excitation. Experimental values are confirmed by numerical simulations using the Khokhlov–Zabolotskaya–Kuznetsov equation on a spatially convergent grid.
 [DOI: 10.1121/1.3298436]

PACS number(s): 43.25.Lj, 43.25.Cb, 43.25.Jh [ROC] Pages: 1208–1217

I. INTRODUCTION

Over the past 20 years, high-frequency diagnostic ultrasound (>10 MHz) has found wide interest in areas such as intravascular ultrasound (IVUS),[1] small animal imaging,[2] and ocular diagnostics.[3,4] The improved spatial resolution at higher frequencies allows for imaging smaller anatomic or pathologic structures, such as atherosclerotic plaques. The implementation of flow imaging at higher frequencies has improved both sensitivity and resolution to permit microvascular imaging.[4,5] There remain, however, limitations in sensitivity due to issues of noise and tissue motion, such as in the context of vasa vasorum imaging in coronary arteries. This has led to investigating the use of ultrasound contrast agents[6] (UCAs) that enhance echo signals from the cardiovascular system, thus increasing the contrast to the surrounding tissue. The majority of ultrasound contrast imaging techniques are most efficient if the imaging frequency of the diagnostic system is approximately equal to the resonance frequency of the bubbles.[7,8] Commercial UCAs have therefore been developed to be primarily resonant within a limited frequency range, which is usually lower than the frequencies used in the aforementioned applications. For example, while contrast-enhanced IVUS and small imaging systems employ imaging frequencies of 15–50 MHz, the major resonance frequency of most conventional UCA bubbles is generally around 1–5 MHz.[9–11] The exception to this is Definity (Lantheus Medical Imaging, North Billerica, MA), which exhibits a dual frequency resonant peak, with a substantial subpopulation of bubbles resonant at higher ultrasound frequencies.[12] As a result, nonlinear (second and subharmonic) imaging approaches at higher frequencies have been shown to be feasible with Definity, either in native or population modified form, or with experimental bubble formulations that have been fabricated to be comprised primarily of bubbles below approximately 2 μm in diameter.[13] Despite the potential of second and subharmonic imaging of small bubble populations, it has yet to be established that these approaches are optimal for higher frequencies, which motivates the investigation of other approaches.

Often it is suggested in the literature that the detection of UCAs can be improved by using two-frequency approaches such as radial modulation[14–17] and nonlinear mixing.[18–21] Due to a nonlinear response of bubbles, the two frequencies would produce sum-and-difference components, which can be detected. For example, two studies exploited the beat frequency f_b of two carrier waves f_1 and f_2, where $f_b \ll f_1, f_2$.[20–22] The beat frequency was scattered by a population of bubbles and detected by a separate low-frequency transducer. Wu *et al.*[19] earlier used a similar technique, where the low-frequency was generated as a cross-product of all frequency contents of a short pulse. Using this method, Wu *et al.* reported a contrast-to-tissue ratio of ±15 dB (concentration was not reported) that the authors addressed fully to the nonlinear properties of bubbles.

However, in the rationale for such dual frequency UCA bubble detection techniques, it is neglected that sufficiently high pressure levels (on the order of 200 kPa and more), such as those used in the experiments, could introduce significant distortion of the pulse due to nonlinear propagation in the medium. More generally, clinical diagnostic ultrasound is regularly applied at pressure levels that introduce nonlinear distortion of the imaging pulse, not only leading to higher harmonics of the fundamental frequency[23,24] but also to low-frequency contents due to the so-called self-demodulation effect.[25–27] Self-demodulation is an effect in which a low-frequency pressure wave develops in a nonlinear medium as a function of the envelope of a high-frequency pulse. The low-frequency wave propagates as if it was generated by a planar low-frequency source and has di-

[a]Present address: Sunnybrook Health Sciences Centre, University of Toronto, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada.
[b]Author to whom correspondence should be addressed. Electronic mail: n.dejong@erasmusmc.nl

rectivity about equal to the high-frequency carrier wave. Hence, the term "parametric array" has often been used in the past to describe such wave generation, particularly in marine applications.[25,28] Furthermore, self-demodulation has been experimentally shown to be able to excite uncoated gas bubbles.[29] The self-demodulation of short pulses was experimentally confirmed using carrier frequencies of 3.5 (Ref. 26) and 10 MHz.[30]

In an accompanying study, for which preliminary results are reported in Ref. 31, the current writers have found that a short (~2 cycles) ultrasound pulse with a carrier frequency of 10 MHz can excite bubbles having a resonance frequency of about 1 MHz by repeating the pulse at a repetition frequency of 1 MHz. In Ref. 31 it was suggested that the self-demodulation effect is the most significant mechanism to convert energy from the 10 MHz carrier wave into the lower bubble resonance frequency. However, no previous reports were found to address this effect in relevant settings for UCAs and high-frequency diagnostic ultrasound. This manuscript therefore characterizes the self-demodulation effect using both numerical simulations and acoustical experiments at frequencies and pressure levels suitable for high-frequency contrast-enhanced ultrasound imaging.

II. METHODS

A. Nonlinear ultrasound propagation

1. Transition pressure level for shock waves

The propagation of ultrasound can be described in three different regimes.[23,32,33] The first is the acoustical limit, which applies for very small amplitudes of the acoustic wave. The propagation is linear in such case. The second is a weakly nonlinear regime, in which waves can interact with other waves and itself; nonlinear distortion gives rise to second and higher harmonics of the fundamental frequency and also to lower frequencies due to the self-demodulation effect of non-continuous waves. The third approximation is the shock-front regime, in which nonlinear response of the medium has deformed an initially sinusoidal wave into sawtooth-like waves including a spatial discontinuity of the pressure. Such spatial discontinuity leads to high nonlinear dissipation of the acoustic energy.[27,32,34,35]

The current study addresses both the weakly nonlinear and shock-wave regimes. The transition pressure level between the two regimes can be estimated by considering the ultrasound field and medium properties. Bacon[34] theoretically derived the pressure in the focal region of a focused transducer for which the shock-front is fully developed. Equation (6) in Ref. 34 is rewritten as

$$p_f = \frac{\sigma_d \rho c^3 \sqrt{G^2-1}}{\omega \beta d \ln(G+\sqrt{G^2-1})}, \tag{1}$$

where p_f is the peak rarefaction pressure in the focus, σ_d is a dimensionless parameter related to the amount of distortion of the pulse[23,36] and the shock-front is fully developed for $\sigma_d = \pi/2$,[34] ρ is the density of the medium (998 kg/m^3 for water), c is the speed of sound (1482 m/s), ω is the driving frequency (rad/s), β is the nonlinear parameter [3.5 (Ref. 23)], and d is the geometrical curvature of the surface approximately equal to the focal distance if G is $\gg 1$. The acoustical gain G describes the focusing strength of a transducer defined by the ratio of the pressure level in the focus to the pressure level at the transducer surface, neglecting attenuation. The gain can be calculated by $G = \omega a^2/2cd$, with a as the aperture radius of the transducer.

Equation (1) assumes a Gaussian axial beam profile and neglects attenuation. Nevertheless it will serve as an estimator for the transition pressure from weakly nonlinear distortion to the full-shock theory throughout this paper.

2. Weakly nonlinear approximation of self-demodulation

As was derived by Berktay[25] in 1965 based on Westervelt's equation, the self-demodulation wave can be described as a function of the pulse envelope $E(t)$ only, i.e., independent of the carrier wave frequency and phase, as long as the time-scale of the envelope is far lower than the carrier frequency. Then, the self-demodulation pressure wave $p_{sd}(t)$ can be approximated by[25]

$$p_{sd}(t) \propto \frac{d^2}{dt^2}|E(t)|^2. \tag{2}$$

The current study exploits a transmitted ultrasound field $p_t(t)$ composed of a carrier wave having sine-wave envelope

$$p_t(t) = A \sin\left(\frac{\omega_{sd}}{2}t\right)\sin(\omega_c t), \tag{3}$$

which when employing Eq. (2) becomes

$$p_{sd}(t) \propto A^2 \omega_{sd}^2 \sin(\omega_{sd} t), \tag{4}$$

where A is the wave amplitude, ω is the angular frequency, "sd" denotes self-demodulation, and c is the carrier wave ($\omega_{sd} \ll \omega_c$). Since $E(t)$ is a sine-wave envelope in Eq. (3), $|E|^2$ has a frequency two times the frequency of $E(t)$. Furthermore, the second time derivative has a magnitude that depends linearly on $\omega_{sd}^2 E^2$. Hence the amplitude of the low-frequency component $p_{sd}(t)$ in Eq. (4) relates quadratically to the envelope frequency.

The transmitted pulse $p_t(t)$ in Eq. (3) is described likely by

$$p_t(t) = \frac{A}{2}\sin(\omega_1 t) + \frac{A}{2}\sin(\omega_2 t), \tag{5}$$

where $\omega_1 = \omega_c - \omega_{sd}/2$ and $\omega_1 = \omega_c + \omega_{sd}/2$. Hence, the frequency spectrum of $p_t(t)$ shows two separated peaks centered around the carrier frequency ω_c, with a difference frequency equal to ω_{sd}, the self-demodulation frequency.

A more thorough analysis of such bi-frequency source signal is found in the Fenlon solution,[32,37] which also addresses the nonlinear interaction between secondary frequency components, i.e., from all sum-and-difference combinations of $q\omega_1 + r\omega_2$, where q and r are integers. However, Eq. (2) provides an easy concept with the limitation that the approximation is only valid for low-level source pressures when secondary effects such as multiple harmonics of the

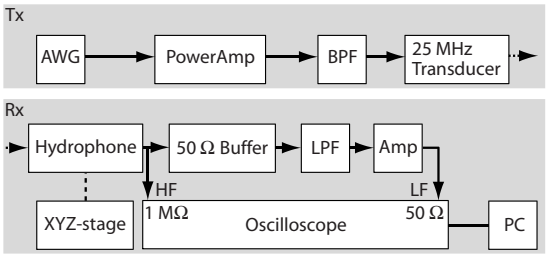


FIG. 1. Block diagram of the experimental setup. AWG is the arbitrary waveform generator, BPF is band-pass filter (8–28 MHz), and LPF is low-pass filter (10.7 MHz).

difference frequency are negligible. At the pressure levels reported in the current study, the secondary effects including shock-fronts will be shown to be significant, motivating the use of a numerical simulation model.

3. Numerical ultrasound propagation model

Simulation models of nonlinear propagation of ultrasound have found wide interest in the past.[23,35,38–42] One common approach for calculating the ultrasound field is the implementation of a parabolic wave-equation, as developed by Khokhlov, Zabolotskaya, and Kuznetsov (KZK).[32] The KZK-equation incorporates diffraction, absorption, and nonlinear propagation. As described by Lee and Hamilton,[38] the equations can be solved in the time-space-domain using a finite difference algorithm. Such approach has been shown to be able to predict the self-demodulation of a short pulse[26,38,43,44] and shock-fronts.[24,35]

In the experiments described in this paper, a strongly focused transducer was used, having an acoustical gain of 19. While the implementation of Lee and Hamilton is strictly limited to gains lower than 10, Hart and Hamilton[39] proposed a simulation grid that is suitable for strongly focused transducers. In Ref. 39 the KZK-equation is solved in frequency domain, while the current authors have chosen to adapt the time-domain numerical algorithm of Lee and Hamilton to work on the converging grid proposed by Hart and Hamilton. Details of the implementation are found in the Appendix. In the simulations the attenuation in water was taken as 25×10^{-3} Np/m MHz^2,[45] i.e., a sound diffusivity of 4×10^{-6} m^2/s (Ref. 32) and a converging-grid parameter δ of 7.[39] The grid resolution close to the transducer surface in the lateral direction was 16 μm (relative resolution of 0.005 of the aperture radius) and the axial stepsize was 38 μm. In focus these were 14 and 29 μm, respectively, showing the convergence of the grid. The temporal resolution was 0.66 ns.

B. Experimental setup

In order to verify the numerical results, an experimental setup was used to measure the actual ultrasound field of a 25 MHz single element transducer (V324, 1 in. focus, Olympus NDT, Waltham, MA). This transducer has an effective diameter of 6.3 mm and 27 mm focal distance, giving an acoustical gain of 19 at 25 MHz. With reference to Fig. 1, an arbitrary waveform generator (AWG520, Tektronix, Beaverton, OR) was connected through a −4 dB precision attenuator (VAT-4+, Mini-Circuits, Brooklyn, NY) to a linear power amplifier (1000 W, 10–86 MHz LPI-10, ENI, Rochester, NY). The output of the power amplifier was band-pass filtered using a custom high-power, 8–28 MHz, fifth-order Butterworth filter to reduce any residual low-frequency content and fed to the transducer. The transducer is characterized to generate a surface pressure of ~1.7 MPa peak-peak when driven by a 234 V peak-peak signal at 25 MHz. The acoustic pressure was measured by a calibrated needle hydrophone (0.075 mm diameter PVDF, Precision Acoustics Ltd., Dorchester, Dorset, UK) immersed in a large water tank. The tank contained tap-water left overnight to obtain room-temperature gas saturated water. The hydrophone was mounted on a computer-controlled precision XYZ stage. Although significant low-frequency content was present in the electric signal at the output of the high-power amplifier due to nonlinear amplification, this was strongly reduced by the band-pass filter at the output (reduction of >40 dB at 1.5 and 3 MHz) and by the reduced efficiency of the transducer at 1.5 and 3 MHz (additional reduction of >20 dB).

The expected low-frequency pressure levels are in the order of −30 to −50 dB below the high-frequency levels, and the receiving electronic circuit had to be extremely linear in order to avoid any nonlinear mixing. Therefore, the signal was split into low-frequency (LF) and high-frequency (HF) tracks. Again with reference to Fig. 1, the hydrophone signal was fed to a first channel of a 1 Gsample/s digital oscilloscope (Tektronix TDS3014B) set to 1 MΩ impedance and 256-averaging mode (HF track). The hydrophone signal was in parallel fed to a custom-built 50 Ω buffer amplifier (buf 602-opamp on a TI DEM-BUF-SO-1A demonstration fixture, Texas Instruments, Dallas, TX), a low-pass filter (10.7 MHz, Mini-Circuits, Brooklyn, NY), and amplified with a 36 dB low-noise amplifier (MiTEQ AU-1189, Hauppauge, NY) (LF track). This signal was averaged on a second channel of the oscilloscope.

The signal levels were carefully checked to be within the linearity-limits of the internal hydrophone amplifier, buffer amplifier, and low-noise amplifier. Also the amplification and attenuation of the components have been carefully characterized over the full range of frequencies employed (1–50 MHz). Final signal to noise ratio values were at least 17 dB at the low-frequency range (1–10 MHz) and 52 dB at the high-frequency range (20–60 MHz).

III. RESULTS

A. General principle

In the numerical simulation a pulse is constructed by summation of sinusoidal signals of 23.5 MHz (f_1) and 26.5 MHz (f_2) (see Fig. 2, left panel). The amplitude of the source pressure is set to 85 kPa, which results in focal pressures in the weakly nonlinear regime. The bottom panel shows the normalized frequency spectrum of the transmitted pulse obtained by the Fourier transform. The propagated pulse (Fig. 2, top middle) shows a rarefaction pressure level (0.7 MPa) smaller than the compression pressure level (1.6 MPa), demonstrating the distortion of the originally symmetric pulse. The spectrum in the bottom-middle panel contains the fun-

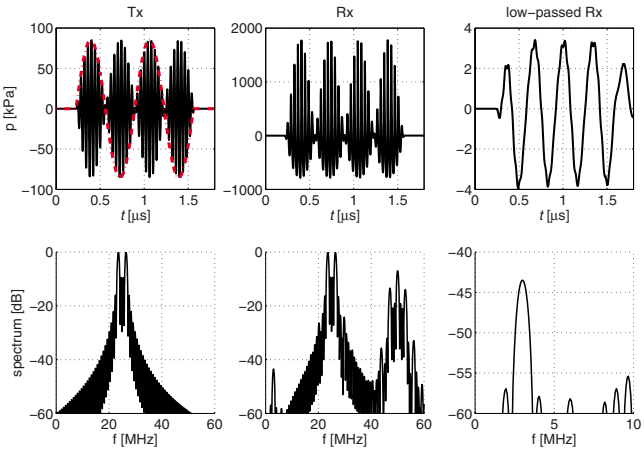


FIG. 2. (Color online) Simulation result of transmitting an ultrasound pulse constructed by the summation of 23.5 and 26.5 MHz waves, with 85 kPa amplitude (upper left), its frequency contents (lower left), and the resulting pulse (middle) and low-pass filtered pulse (right) in focus. The conversion of energy to lower frequency is caused by the self-demodulation effect.

damental frequencies f_1 and f_2, their second harmonic frequencies $2f_1$ and $2f_2$, and the sum-and-difference components of 50 and 3.0 MHz.

The pulse in focus is low-pass filtered (sixth-order Butterworth, −6 dB cut-off point at 6 MHz) and shown in Fig. 2 (top right panel). The amplitude of the low-frequency content is about 3 kPa. The corresponding spectral densities at the low-frequency are shown in the bottom panel. Note that the filtering operation is linear, and the low-frequency content is a "physical" wave that can autonomously interact with a bubble, even if the high-frequency carrier wave is attenuated in the focus.

Similarly, Fig. 3 shows the spectra of focal pulses with respective envelopes of 0 (i.e., no modulation), 1.5, and 3

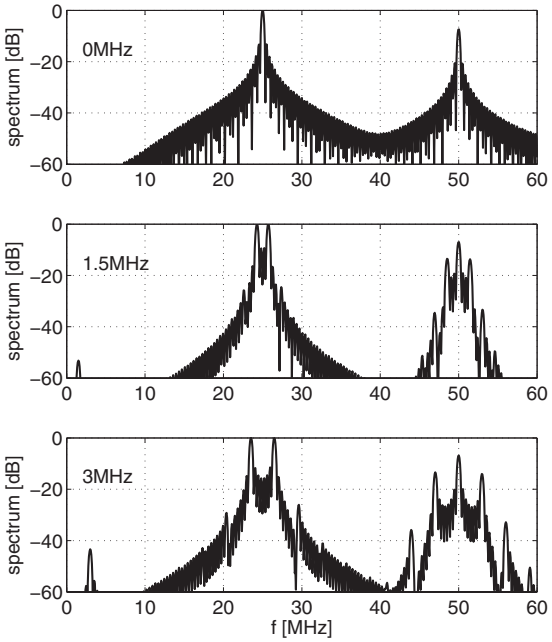


FIG. 3. Fourier spectra of weakly nonlinearly propagated pulses having center frequency of 25 MHz and source pressure amplitude of 85 kPa. The envelope frequencies are 0, 1.5, and 3 MHz.

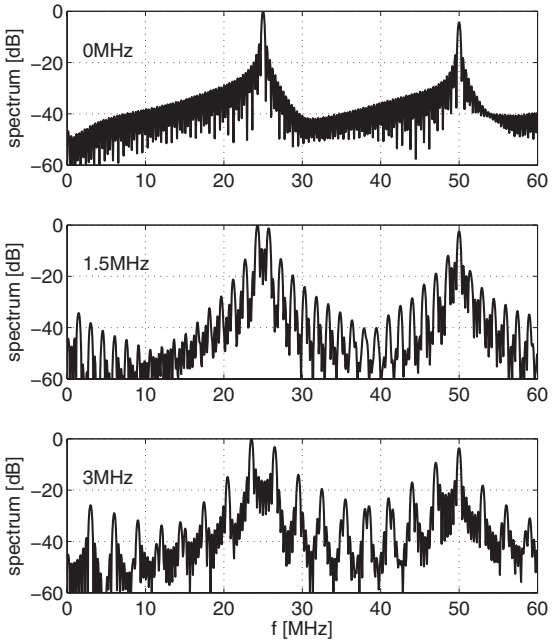


FIG. 4. Same parameter values as in Fig. 3, here with 850 kPa source pressure.

MHz. The increase in difference frequency f_1-f_2 leads to higher level of the self-demodulation frequency, as expected from application of Eq. (4). The time-domain pulses showed amplitudes of 1.1 kPa for the 1.5 MHz-envelope pulse and 5.7 kPa for the 3 MHz-envelope pulse. Such increase with a factor of 5 with double envelope frequency is roughly predicted by Eq. (4). With respect to the tone burst, a self-demodulation signal will appear, depending on the second derivative of the envelope squared [see Eq. (2)], but the level of this signal is outside the displayed dynamic range.

When setting the source pressure amplitude to 850 kPa, the threshold of 1.7 MPa of focal pressure predicted by Eq. (1) is exceeded. The pulse in focus has formed a full- shock-front, with a peak-negative value of 1.7 MPa and peak-positive value of up to 7 MPa. The spectra in Fig. 4 show many more sum-and-difference components, originating from interactions of both primary and secondary waves. The pulse contains strong second and higher harmonic contents of the low-frequency wave; for example, with the 3 MHz-envelope frequency, low-frequency waves with a frequency of 6 and 9 MHz are visible. Furthermore, the high-frequency contents show many side-bands with interspacing frequencies equal to the envelope frequency. Also, cross-products between f_1, f_2, and f_{sd} show up, at, e.g., 20.5 and 29.5 MHz, for a 3 MHz-envelope. Low-pass filtered waves had time-domain amplitudes of 45 and 130 kPa for the 1.5 and 3.0 MHz envelopes respectively, indicating a pressure increase by roughly a factor of 3 when the envelope frequency is doubled. In comparing this to the weakly nonlinear example, this shows that the formation of the shock-front reduces the squared dependency of the self-demodulation amplitude on the envelope frequency.

B. Transmit pressure dependency

Figure 5 shows the influence of the source pressure amplitude on the frequency content of the pulse at focus. For

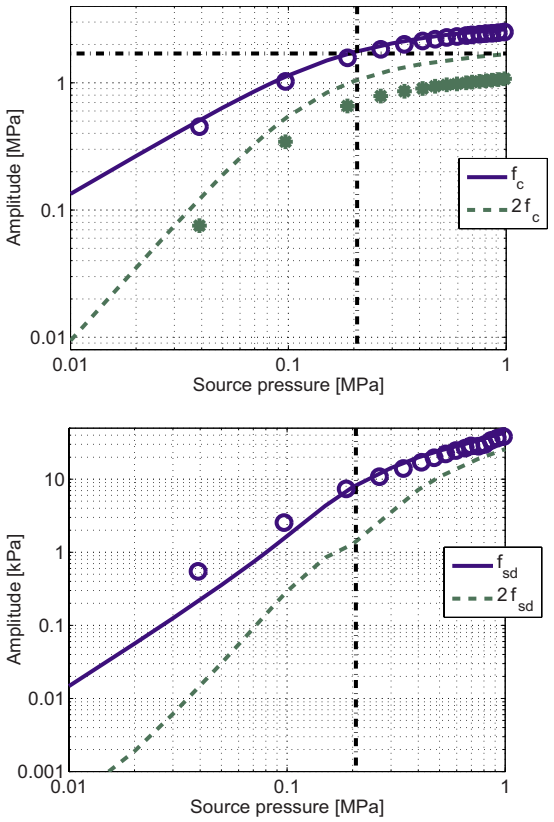


FIG. 5. (Color online) High-frequency (top) and low-frequency (bottom) pressure levels in the focus as a function of the source pressure. Circles denote experimental values. Carrier frequency f_c was 25 MHz and self-demodulation frequency f_{sd} was 1.5 MHz. The vertical line marks the full-shock pressure threshold of 1.7 MPa, as predicted by Eq. (1) for this transducer.

amplitudes lower than approximately 200 kPa, corresponding to a full-shock threshold of 1.7 MPa pressure in focus [Eq. (1), dashed-dotted lines in Fig. 5], the high-frequency f_c component has a linear relationship with source amplitude, whereas the low-frequency amplitude f_{sd} depends quadratically on the transmit pressure, as expected in the weakly nonlinear regime. The second harmonic of the low-frequency content ($2f_{sd}$) depends approximately cubically on the source level. This is unexpected since the bi-frequency solution of Fenlon predicts a power-of-4 dependency.[37] The difference will be further addressed in Sec. IV.

Above 200 kPa source pressure, the fundamental f_c component in the focus starts to deviate from a linear dependency, showing a flattening toward the acoustic saturation limit.[32,41] This is consistent with the expectation of nonlinear increase in dissipation associated with the presence of a shock-front.[41] Furthermore, the low-frequency (f_{sd}) amplitude starts to deviate from squared dependency to an almost linear relation, which is different from propagation in the weakly nonlinear regime.

Note that although the theoretical acoustical gain of this transducer at 25 MHz is 19, the actual ratio between source pressure and focal pressure for low amplitudes is only 12.5 [see Fig. 5(a), e.g., source pressure: 10 kPa, focal pressure: 125 kPa]. The difference of 52% (3.6 dB) can be fully attributed to the attenuation in water, which is about 2.2

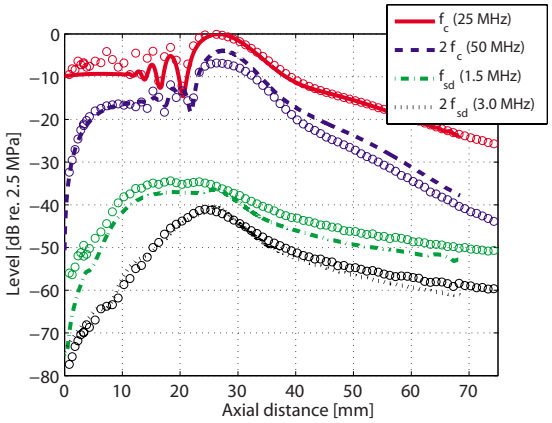


FIG. 6. (Color online) Simulated (lines) and experimental (circles) axial beam profiles at the fundamental frequency (25 MHz), second harmonic (50 MHz), self-demodulation frequency (1.5 MHz), and double self-demodulation frequency (3 MHz) as a function of distance from the transducer. Source amplitude is 850 kPa.

$\times 10^{-3}$ dB cm^{-1} MHz^{-2}.[45] At 25 MHz and a focal distance of 27 mm, the ultrasound is expected to be attenuated by 3.7 dB.

Experimentally obtained values are shown by dots in Fig. 5. The measured values of f_c and f_{sd} agree within 10% of the numerical predictions. The second harmonic amplitude at $2f_c$ is about 40% (3 dB) lower than the predicted amplitudes throughout the full range.

C. Beam profiles

Figure 6 shows the predicted axial beam profile of the transducer in the full-shock regime. The envelope frequency is set to generate 1.5 MHz self-demodulation waves and the pressure at the transducer surface is 850 kPa in amplitude, similar to the experimental value. Pulses obtained along the beam axis are digitally band-pass filtered at the carrier wave frequency (20–30 MHz, −6 dB bandwidth, Butterworth fourth-order) at the second harmonic frequency (40–60 MHz), at the self-demodulation frequency (0.75–2.25 MHz), and at twice the self-demodulation frequency (2.25–3.75 MHz). The plotted amplitudes are obtained in the time-domain after Hilbert transforming the filtered pulses.

At the carrier frequency f_c and its second harmonic, the highest amplitudes appear at 27 mm, which is the focal region. The self-demodulation wave begins to plateau at a 15 mm distance from the transducer, reaching a peak amplitude of −35 dB before decreasing after 25 mm. The amplitude of the 3 MHz wave, twice the self-demodulation frequency, is up to −40 dB in focus. Since 0 dB corresponds to 2.5 MPa in this figure, the amplitudes of the waves at 1.5 and 3 MHz are 45 and 25 kPa, respectively.

Figure 6 also shows the experimentally obtained amplitudes. The hydrophone was positioned along the beam axis at incremental distances from the transducer of 1–75 mm, in steps of 1 mm. The amplitude levels are obtained from the recorded pulses with the same filter characteristics as in the simulation and compensated for the hydrophone sensitivity. At 1 mm from the transducer, the amplitude of the self-demodulation pulse is over 50 dB smaller than the funda-

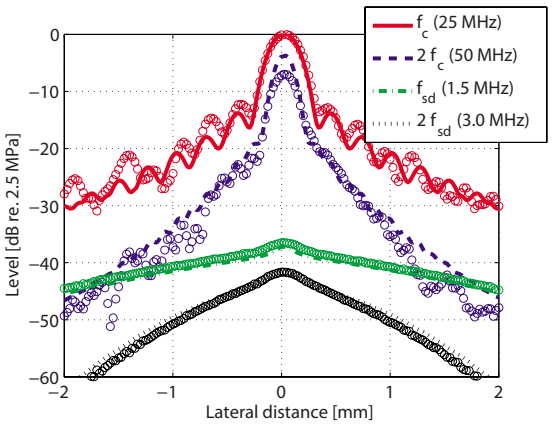


FIG. 7. (Color online) Simulated (lines) and experimental (circles) lateral beam profiles in focus (27 mm from transducer).

mental level, but achieves a maximum of −35 dB in the region before the focus. The normalization value was equal to that of the simulation. Note that the level at the second harmonic frequency matches the predicted values except for the focal region, where a difference of 3 dB is observed (cf. Fig. 5).

The lateral beam profile in the focal region (27 mm from the transducer) is plotted in Fig. 7. The −6 dB width of the fundamental frequency beam is 0.4 mm, while the self-demodulation beam measures approximately 3 mm. The high-frequency second harmonic beam at 50 MHz has about 0.3 mm width, and the low-frequency second harmonic beam at 3 MHz has 1.4 mm width.

The experimental lateral beam shown in Fig. 7 is very similar in shape and amplitude to the predicted beam. The fundamental-frequency side lobes show a slightly different spatial frequency, which could be attributed to a slight error in the estimation of the transducer geometry used for the simulation. Furthermore, the second harmonic levels are well predicted except for within the focal region and for $z \geq 0.7$ mm.

D. Influence of focusing and medium

Based on the simulation model, some additional comments can be made regarding the application of the self-demodulation technique.

First, focusing generally increases the pressure levels and therefore the self-demodulation pressure. For example, when modeling a focal distance of 13.5 mm (acoustic gain of 40) instead of 27 mm, with source pressure of 850 kPa, the self-demodulation pressure increases from 45 to 170 kPa. The change in focal distance also increases the primary wave amplitude in focus, and for comparison, reducing the source pressure in the latter configuration to 425 kPa gives a 92 kPa self-demodulation wave. Note that in such comparison the self-demodulation wave depends almost linearly on the source pressure, and not quadratic, as expected from Eq. (4), since nonlinear shock-wave dissipation is significant. As another example, taking a flat surface (i.e., unfocused) instead of curved one, the self-demodulation level is 4 kPa. Compared to the focused geometry the conversion is much lower.

TABLE I. Simulation results with two transducers models, in water and tissue. Carrier frequency is 25 MHz and envelope frequency is 1.5 MHz.

Aperture diameter (mm)	Focal distance (mm)	Acoustical gain	Source pressure (MPa)	Amplitude of self-demodulation wave (kPa)	
				Water	Tissue
6	12.7	36	0.5	100	4
			5	800	250
0.9	Flat	2	1.5	7	3
			5	60	30

This can be attributed to the lower acoustical gain and high attenuation in water at 25 MHz, giving lower fundamental-frequency pressure levels.

Second, the application of the self-demodulation technique *in vivo* implies a different propagation medium. Tissue generally has a higher nonlinearity value than water, but also higher attenuation. For example, liver tissue has a nonlinearity β of 4.4 compared to 3.5 of water.[45] The attenuation is about 20 dB/cm at 25 MHz, while water has attenuation of about 1.4 dB/cm at 25 MHz.[45] Combining tissue values with the focused transducer model gives a maximum self-demodulation pressure level of 6 kPa (peak rarefaction amplitude of 1.1 MPa at the 25 MHz primary field), which is low for contrast excitation applications.[46] In clinical applications it would be better to use strongly focused transducers. For example, a 25 MHz probe of the commercially available high-frequency scanner Vevo-770 (RMV 707B, VisualSonics, Toronto, Canada) has a diameter of 6 mm and focal length of 12.7 mm (gain of 36). The acoustic pressure close to the transducer was 0.5 MPa measured with the hydrophone setup. Inserting such model in the simulation with liver tissue, the self-demodulation pressure is 4 kPa. By artificially increasing the source pressure to 5 MPa, which might be achieved in future using more efficient transducers,[47] a self-demodulation value of 250 kPa can be obtained (see Table I). The peak rarefaction pressure value of 25 MHz wave is 4.5 MPa. Note that the high attenuation has resisted the formation of shock waves, giving an almost squared relation of the self-demodulation with source pressure in tissue; this relation could help evaluate the applicability of high-frequency systems for self-demodulation techniques.

Third, the combination of the self-demodulation technique with IVUS transducers[1] might be beneficial for contrast agent excitation. Such catheter based transducers are limited in size to about 0.9 mm in diameter, and the source pressure can reach higher values, currently up to at least 1.5 MPa. For this application, the simulated self-demodulation pressures will not exceed 3 kPa in tissue. We estimate that it would be necessary to increase IVUS source pressures by a factor of 3 in order to generate a self-demodulation signal sufficiently strong to excite contrast agents (see Table I).

IV. DISCUSSION AND CONCLUSIONS

A. General principle

Figures 2–5 show the general results of the self-demodulation technique. The full-shock threshold level is

well predicted by Eq. (1), even in the case of strong attenuation at such high carrier frequencies (Fig. 5). A focal pressure far below this threshold shows self-demodulation tendency described by Berktay[25] [cf. Eq. (2)], but a pressure close to the threshold resulted in significant spectral content at many distinct frequencies. In particular, the second and higher harmonic frequencies of the low-frequency wave are clearly visible in Fig. 4. The origin of these signals is the interaction of primary and/or secondary frequency components. For example, a 3 MHz wave could originate from the interaction of 22.75 (f_1-f_{sd}) and 25.75 MHz (f_2), or secondary frequencies only, for example, 48.5 ($2f_1$) and 51.5 MHz ($2f_2$). One way to look at the higher difference frequency is based on Berktay's approximation [Eq. (2)]. Because the envelope shown in Fig. 2 (middle panel) is not a sine-wave anymore, but rather a sum of sine-waves having multiple-integer frequencies of the fundamental, Eq. (2) will also predict multiple-integers of the self-demodulation frequency.

The third-power relation between the amplitude of the second difference frequency ($2f_{sd}$) and the amplitude of the source pressure is not predicted by nonlinear plane-wave propagation theory. Fenlon's solution[32,37] predicts that the amplitude of the second difference frequency should increase with the source pressure to the fourth-power, rather than the third-power increase found in the current simulations. Furthermore, we employed additional calculations in which the diffraction step was removed from the KZK simulations. This basically yields a numerical implementation of the plane-wave Burgers equation in time-domain.[32] When the other simulation parameters were kept equal to those described in the current paper, a fourth-power increase was found, like predicted by the Fenlon solution. Variation in the attenuation values in this plane-wave simulation did not change the result either. Hence, the deviation from the fourth-power dependency is not caused by the attenuation or numerical implementation. It can therefore be suggested that the focusing of the current sound field is responsible for the observed third-power dependency.

Another striking feature is seen in the lateral beam profiles at 1.5 and 3 MHz (Fig. 7). The −6 dB beam width of the low-frequency wave is 3 mm, which equals 3λ (wavelength $\lambda \approx 1$ mm at 1.5 MHz). The 3 MHz wave has −6 dB beam of 1.4 mm, again equal to about 3λ ($\lambda \approx 0.5$ mm at 3 MHz). The absolute beam width decreases with the harmonic number, as was earlier reported,[27] and the equality of the relative beam widths implies equal sources of the secondary waves, as long as the phenomenon is diffraction limited. Also note that the carrier beam width in focus is only 0.4 mm, much smaller than that of the self-demodulation wave, but is much wider in the pre-focal region where the self-demodulation wave is building up.

The −6 dB beam width of 3 mm (Fig. 7) could be somewhat disappointing for imaging applications since the parametric approach has been used in the past to generate narrow low-frequency beams with relatively small transducers. However, conventionally producing such beam width at 27 mm distance from a 6.3 mm transducer is impossible at 1.5 MHz (from KZK simulation: 6 mm beam width at 27 mm distance).[25] Also, the second harmonic beam width of such transducer transmitting a 0.75MHz ultrasound field still produces larger beam width (4 mm). Therefore, the self-demodulation technique provides narrowest beam. Apart from the lateral beam width, the axial focal region is smaller and more pronounced than with conventional 1.5 MHz ultrasound field generation.[48]

When comparing the simulation results of the beam profiles to the experimental ones (Figs. 6 and 7), in general, very similar levels occur for all four frequency components (except for the focal region; see below). This finding has two major implications. First, it implies that the simulations using the KZK-equation predicts correct attenuation and diffraction for the given transducer model. Second, it also correctly predicts the nonlinear distortion of the pulse. These are not trivial results since the parabolic assumption in the wave propagation is challenged by the geometries employed in this study. However, Naze Tjøtta *et al.*[48] derived that the parabolic approximation is valid as long as the ratio of focal distance d and aperture radius a are larger than 3.5, which is the case in the current study ($d/a \sim 8$).

The $2f_c$ level in focus is predicted by the model to be −4 dB compared to fundamental level, while the measured value is at −7 dB; this difference is also seen in Fig. 5. A −4 dB second harmonic level is perhaps surprising since nonlinear propagation theory predicts an absolute maximum of −6 dB.[23] The current writers can only hypothesize that the error originates from an underestimation of the shock-wave dissipation since the shock-wave is most prominent in the focal region.

B. Implications

The role of attenuation is twofold in the self-demodulation process. In general terms the self-demodulation process merits a highly absorptive medium to isolate the low-frequency components from the high-frequency primary waves. However, attenuation also reduces the pressure level, which, in turn, could affect the nonlinear conversion. Focusing of the beam produces a 40-fold increase in the self-demodulation amplitude in water, as shown with the additional simulations, because of the higher gain but also because of less attenuation since the distance to the focal region is reduced. Note that Naze Tjøtta *et al.*[48] earlier stated that focusing has only limited effect on the self-demodulation amplitude, but their finding was based on zero-absorption calculations.

The self-demodulation technique allows for engineering pulses, e.g., without any transient head or tail (cf. Fig. 2), or for generating very wide-band low-frequency signals.[25,29,40] This can be advantageous over conventional linear generation of pulses in an experimental setup with a possibly limited bandwidth. To produce low-frequency wide-band signals, a high-frequency carrier wave with Gaussian envelope suffices.[25,29,40] In fact, by carefully applying Berktay's approximation on a Gaussian-enveloped pulse, it can be shown that the center frequency of the self-demodulated wave is equal to the −3 dB absolute bandwidth of the transmission pulse.[25] This rule of thumb could help in analyzing the frequency spectrum of a nonlinearly propagating acoustic pulse.

Furthermore, the technique provides acoustic amplitudes suitable for contrast microbubble excitation. This has two major implications. First, the effect can, for example, be used in spectroscopy measurements, as shown by the current authors.[31] The high pressure amplitudes of the primary field in water could give rise to microbubble destruction that should be avoided. An acoustic filter[40] or absorptive medium such as tissue could reduce the level of the primary waves. A pressure value of 1.1 MPa in tissue at 25 MHz, as reported in Sec. III, is close to the microbubble destruction level at 20 MHz (Ref. 49) but below that of 30 MHz.[13] However, the approach is most suitable for imaging larger bubbles (i.e., resonant at the low frequencies), and these will be far less susceptible to being destroyed by high-frequency ultrasound than smaller bubbles that are closer to resonance. Second, the finding might be important for the works by, e.g., Wu *et al.*[19] and Chen *et al.*,[20] who used a semi-dual frequency technique to increase the detection signals of contrast bubbles. In the rationale all low-frequency contents in the echoes are suggested to be generated by nonlinear behavior of microbubbles. The current study, however, shows that nonlinear propagation cannot be *a priori* neglected. In contrast detection applications where the main low-frequency generation mechanism is nonlinear propagation, the low-frequency content will also be present in tissue-scattered signals, and the agents merely increase contrast by linear scattering. Hydrophone experiments or numerical propagation simulations showing the exact frequency content of the acoustic excitation pulse could elucidate the relative significance of nonlinear propagation compared to nonlinear scattering of bubbles.

C. Concluding remarks

This paper presents numerical and experimental results of the self-demodulation effect, using ultrasound conditions that are relevant for high-frequency medical imaging. The mechanism provides sufficient acoustic pressure in water, 45 kPa, to excite an UCA microbubble.[46,50] Although the self-demodulation technique has been previously used to investigate uncoated gas bubbles in water, see, e.g, Refs. 29 and 40, no reports have been found to address the possibility of its use in combination with UCAs.

The poor efficiency of the self-demodulation mechanism is a disadvantage. In the reported experiments, the source pressure is 850 kPa, giving 45 kPa low-frequency amplitude. Tissue attenuation decreases the self-demodulation pressure dramatically (Table I), which can be overcome by focusing and/or increased transmit amplitudes. Another solution is a reduction in carrier frequency to the range 10–20 MHz. Such range is currently used for small animal imaging and carotid artery scanning and might therefore be easily implemented in current clinical systems.

ACKNOWLEDGMENTS

The authors wish to thank Dr. V. F. Humphrey and Dr. J. Huijssen for the very helpful discussions on the self-demodulation technique.

APPENDIX: KZK ON A CONVERGING/DIVERGING GRID

Hart and Hamilton described a numerical implementation of the KZK-equation for strongly focused spherical transducers accounting for absorption, diffraction, and nonlinear propagation.[39,48] The spatial numerical grid converges from the transducer to the focal region and diverges afterward, thus having highest resolution in the focal region where high spatial variations in the pressure are expected. The equations developed in Ref. 39 solve the KZK-equation in the frequency domain, while the current writers have opted for an implementation in the time-domain, analogous to the derivation by Lee and Hamilton.[38] Such an implementation in the time-domain is computationally more effective when acoustic shock-fronts are expected compared to the frequency-domain solution.[51] However, the numerical calculation scheme in Ref. 51 has to be adapted to suit the converging/diverging grid.

The reader is referred to Ref. 51 for a full description of the numerical implementation, and Refs. 39 and 48 for the theory on the converging/diverging grid. The KZK-equation in reduced units for the converging/diverging grid reads[39] as

$$\left((\sigma \pm \delta)^2 \frac{\partial^2}{\partial \bar{\tau}\, \partial \sigma} - \frac{1}{4G}\bar{\nabla}_{\perp}^2 - (\sigma \pm \delta)^2 A \frac{\partial^3}{\partial \bar{\tau}^3}\right)\bar{P} = (\sigma \pm \delta)\frac{B}{2}\frac{\partial^2 \bar{P}^2}{\partial \bar{\tau}^2}, \tag{A1}$$

with σ (≥ -1) the axial distance to the transducer, δ (>1) the parameter that governs the convergence of the grid, $\bar{\tau}$ the retarded time (≥ 0), G the acoustic gain, and $\bar{\nabla}_{\perp}^2$ the two-dimensional Laplace operator with respect to $\bar{u}$ (≥ 0), which is the lateral distance to the beam axis. The parameters A and B characterize the absorption and source amplitude, respectively.[39] The pressure signal is denoted by $\bar{P}$. The value of σ equals -1 at the transducer surface and 0 in the geometrical focus of the transducer. The $-$ sign in the denominator is chosen if σ is negative and + elsewhere.[39]

Compared to the time-domain code description in Ref. 51, diffraction and nonlinear propagation have to be adapted for the current grid; attenuation is unchanged. In Ref. 51, Eqs. (3.30)–(3.32) and (3.40–3.42) express the numerical implementation of the diffraction term. These are unchanged. However, the value of R defined in Eqs. (3.33) and (3.43) becomes

$$R = \frac{\Delta\bar{\tau}(\Delta\sigma)_k}{(\sigma_{k+1} \pm \delta)^2 (\Delta\bar{u})^2}, \tag{A2}$$

with k the index associated with finite steps of σ.

Calculation of the nonlinear propagation term is less straightforward due to the change in sign before δ at $\sigma=0$. Equation (3.56) in Ref. 51 is replaced by

$$\frac{\partial \bar{P}}{\partial \sigma} = \frac{N\bar{P}}{(\sigma \pm \delta)}\frac{\partial \bar{P}}{\partial \bar{\tau}}, \tag{A3}$$

with N the parameter of nonlinearity. Equation (A3) is solved using a stretched coordinate ζ,[51]

$$\frac{d\varsigma}{d\sigma}\frac{\partial \bar{P}}{\partial \varsigma} = \frac{N\bar{P}}{(\sigma \pm \delta)}\frac{\partial \bar{P}}{\partial \bar{\tau}}. \tag{A4}$$

With a definition of

$$\frac{d\varsigma}{d\sigma} = \frac{1}{(\sigma \pm \delta)}, \tag{A5}$$

Equation (A4) is reduced to

$$\frac{\partial \bar{P}}{\partial \varsigma} = N\bar{P}\frac{\partial \bar{P}}{\partial \bar{\tau}}, \tag{A6}$$

and Eq. (A5) is rewritten as

$$\varsigma = \int_{-1}^{\sigma} \frac{d\sigma'}{(\sigma' \pm \delta)}, \tag{A7}$$

where the integration limits are dictated by the boundary condition $\zeta=0$ when $\sigma=-1$, and the $-$ sign should be chosen when σ', the integration variable, is negative.

Tedious but straightforward evaluation of Eq. (A7) for all $\sigma<0$ and $\sigma\geq 0$ results in

$$\varsigma = \int_{-1}^{\sigma} \frac{d\sigma'}{(\sigma' \pm \delta)} = \ln\left(\frac{|\sigma| + \delta}{\delta + 1}\right). \tag{A8}$$

Following Lee, an exact implicit solution of Eq. (A6) is of the format

$$\bar{P}(\sigma, \bar{\tau}) = F(\bar{\tau} + N\bar{P}\varsigma), \tag{A9}$$

and at the subsequent spatial step, combining Eqs. (A9) and (A8),

$$\bar{P}(\sigma + \Delta\sigma, \bar{\tau}) = F\left(\bar{\tau} + N\bar{P}\ln\left[\frac{|\sigma| + \Delta\sigma + \delta}{\delta + 1}\right]\right). \tag{A10}$$

Rewriting the argument of the natural logarithm in Eq. (A10) and combining with Eq. (A9) yield the forward stepping algorithm

$$\bar{P}(\sigma + \Delta\sigma, \bar{\tau}) = \bar{P}\left(\sigma, \bar{\tau} + N\bar{P}\ln\left[1 + \frac{\Delta\sigma}{|\sigma| + \delta}\right]\right). \tag{A11}$$

The numerical implementation of the nonlinear differential equation is thus [compared with Eq. (3.71) in Ref. 51]

$$P_{i,j}^{k+1} = \frac{P_{i,j}^{k}}{1 - N[(P_{i+1,j}^{k} - P_{i,j}^{k})/\Delta\bar{\tau}]\ln\left[\dfrac{1 + (\Delta\sigma)_k}{(|\sigma_k| + \delta)}\right]} \quad \text{for } P_{i,j}^{k} \geq 0. \tag{A12}$$

For $P_{i,j}^{k}<0$ the pressure difference takes the form of $(P_{i,j}^{k} - P_{i-1,j}^{k})$, analogous to Lee's implementation.

[1] M. E. Frijlink, D. E. Goertz, H. J. Vos, E. Tesselaar, G. Blacquière, A. Gisolf, R. Krams, and A. F. W. van der Steen, "Harmonic intravascular ultrasound imaging with a dual-frequency catheter," Ultrasound Med. Biol. **32**, 1649–1654 (2006).

[2] D. E. Goertz, M. E. Frijlink, D. Tempel, V. Bhagwandas, A. Gisolf, R. Krams, N. de Jong, and A. F. W. van der Steen, "Subharmonic contrast intravascular ultrasound for vasa vasorum imaging," Ultrasound Med. Biol. **33**, 1859–1872 (2007).

[3] F. S. Foster, C. J. Pavlin, K. A. Harasiewicz, D. A. Christopher, and D. H. Turnbull, "Advances in ultrasound biomicroscopy," Ultrasound Med. Biol. **26**, 1–27 (2000).

[4] D. E. Kruse, R. H. Silverman, R. J. Fornaris, D. J. Coleman, and K. W. Ferrara, "A swept-scanning mode for estimation of blood velocity in the microvasculature," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1437–1440 (1998).

[5] D. E. Goertz, D. A. Christopher, J. L. Yu, R. S. Kerbel, P. N. Burns, and F. S. Foster, "High-frequency color flow imaging of the microcirculation," Ultrasound Med. Biol. **26**, 63–71 (2000).

[6] D. E. Goertz, A. Needles, P. N. Burns, and F. S. Foster, "High-frequency, nonlinear flow imaging of microbubble contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 495–502 (2005).

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

[8] N. de Jong, A. Bouakaz, and P. Frinking, "Basic acoustic properties of microbubbles," Echocardiogr. **19**, 229–240 (2002).

[9] J.-M. Gorce, M. Arditi, and M. Schneider, "Influence of bubble size distribution on the echogenicity of ultrasound contrast agents—A study of SonoVue," Invest. Radiol. **35**, 661–671 (2000).

[10] W. T. Shi and F. Forsberg, "Ultrasonic characterization of the nonlinear properties of contrast microbubbles," Ultrasound Med. Biol. **26**, 93–104 (2000).

[11] K. Sarkar, W. T. Shi, D. Chatterjee, and F. Forsberg, "Characterization of ultrasound contrast microbubbles using in vitro experiments and viscous and viscoelastic interface models for encapsulation," J. Acoust. Soc. Am. **118**, 539–550 (2005).

[12] D. E. Goertz, N. de Jong, and A. F. W. van der Steen, "Attenuation and size distribution measurements of definity and manipulated definity populations," Ultrasound Med. Biol. **33**, 1376–1388 (2007).

[13] D. E. Goertz, M. E. Frijlink, N. de Jong, and A. F. W. van der Steen, "High frequency nonlinear scattering from a micrometer to submicrometer sized lipid encapsulated contrast agent," Ultrasound Med. Biol. **32**, 569–577 (2006).

[14] C. X. Deng, F. L. Lizzi, A. Kalisz, A. Rosado, R. H. Silverman, and D. J. Coleman, "Study of ultrasonic contrast agents using a dual-frequency band technique," Ultrasound Med. Biol. **26**, 819–831 (2000).

[15] A. Bouakaz, M. Versluis, J. Borsboom, and N. de Jong, "Radial modulation of microbubbles for ultrasound contrast imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 2283–2290 (2007).

[16] E. Chérin, J. Brown, S.-E. Måsøy, H. H. Shariff, R. Karshafian, R. Williams, P. N. Burns, and F. S. Foster, "Radial modulation imaging of microbubble contrast agents at high frequency," Ultrasound Med. Biol. **34**, 949–962 (2008).

[17] S.-E. Måsøy, Ø. Standal, P. Näsholm, T. F. Johansen, B. Angelsen, and R. Hansen, "SURF imaging: In vivo demonstration of an ultrasound contrast agent detection technique," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1112–1121 (2008).

[18] D. Cathignol, J. Y. Chapelon, V. L. Newhouse, and P. M. Shankar, "Bubble sizing with high spatial resolution," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 30–37 (1990).

[19] C.-Y. Wu, J. Tsao, and Y.-H. Chou, "An ultrasonic microbubble semi-intermodulated imaging technique," Ultrasound Med. Biol. **31**, 1199–1210 (2005).

[20] S. Chen, R. Kinnick, J. F. Greenleaf, and M. Fatemi, "Difference frequency and its harmonic emitted by microbubbles under dual frequency excitation," Ultrasonics **44**, e123–e126 (2006).

[21] C.-K. Yeh, S.-Y. Su, C.-C. Shen, and M.-L. Li, "Dual high-frequency difference excitation for contrast detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 2164–2176 (2008).

[22] C.-Y. Wu, W.-H. Chao, Y.-H. Chou, and Y. Watanabe, "Analyses of characteristics of semi-intermodulated imaging based on multi-bifrequency approximation," Jpn. J. Appl. Phys., Part 1 **46**, 4841–4846 (2007).

[23] F. A. Duck, "Nonlinear acoustics in diagnostic ultrasound," Ultrasound Med. Biol. **28**, 1–18 (2002).

[24] V. F. Humphrey, "Nonlinear propagation in ultrasonic fields: Measurements, modelling and harmonic imaging," Ultrasonics **38**, 267–272 (2000).

[25] H. O. Berktay, "Possible exploitation of non-linear acoustics in underwater transmitting applications," J. Sound Vib. **2**, 435–461 (1965).

[26] M. A. Averkiou, Y.-S. Lee, and M. F. Hamilton, "Self-demodulation of amplitude-modulated and frequency-modulated pulses in a thermoviscous fluid," J. Acoust. Soc. Am. **94**, 2876–2883 (1993).

[27] J. Naze Tjøtta, S. Tjøtta, and E. H. Vefring, "Propagation and interaction of two collinear finite amplitude sound beams," J. Acoust. Soc. Am. **88**,

2859–2870 (1990).
[28]L. A. Ostrovsky, S. N. Gurbatov, and J. N. Didenkulov, "Nonlinear acoustics in Nizhni Novgorod," Acoust. Phys. **51**, 114–127 (2005).
[29]N. G. Pace, A. Cowley, and A. M. Campbell, "Short pulse acoustic excitation of microbubbles," J. Acoust. Soc. Am. **102**, 1474–1479 (1997).
[30]M. Rénier, C. Barrière, and D. Royer, "Optical measurements of the self-demodulated displacement and its interpretation in terms of radiation pressure," J. Acoust. Soc. Am. **121**, 3341–3348 (2007).
[31]H. J. Vos, D. E. Goertz, and N. de Jong, "Pulse repetition rate excitation of contrast agents," in Proceedings of the IEEE Ultrasonics Symposium, Vancouver, Canada (2006).
[32]M. F. Hamilton and C. L. Morfey, "Model equations," *Nonlinear Acoustics* (Academic, San Diego, CA, 1998), Chap. 3, pp. 41–63.
[33]T. G. Leighton, "Review—What is ultrasound?," Prog. Biophys. Mol. Biol. **93**, 3–83 (2007).
[34]D. R. Bacon, "Finite amplitude distortion of the pulsed fields used in diagnostic ultrasound," Ultrasound Med. Biol. **10**, 189–195 (1984).
[35]O. V. Rudenko, "Nonlinear sawtooth-shaped waves," Phys. Usp. **38**, 965–989 (1995).
[36]F. A. Duck and M. A. Perkins, "Amplitude-dependent losses in ultrasound exposure measurement," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 232–241 (1988).
[37]F. H. Fenlon, "An extension of the Bessel-Fubini series for a multiple frequency CW acoustic source of finite amplitude," J. Acoust. Soc. Am. **51**, 284–289 (1972).
[38]Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[39]T. S. Hart and M. F. Hamilton, "Nonlinear effects in focused sound beams," J. Acoust. Soc. Am. **84**, 1488–1496 (1988).
[40]V. F. Humphrey, "Non-linear acoustics as a laboratory tool," Non-linear acoustics: A tutorial meeting, Proceedings, Institute of Acoustics (UK), Vol. 14, pp. 99–113 (1992).
[41]D. T. Blackstock, "Thermoviscous attenuation of plane, periodic, finite-amplitude sound waves," J. Acoust. Soc. Am. **36**, 534–542 (1964).
[42]J. Tavakkoli, D. Cathignol, R. Souchon, and O. A. Sapozhnikov, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[43]M. A. Averkiou and M. F. Hamilton, "Nonlinear distortion of short pulses radiated by plane and focused circular pistons," J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[44]S. Callé, J. P. Remenieras, O. Bou Matar, and F. Patat, "Presence of nonlinear interference effects as a source of low frequency excitation force in vibro-acoustography," Ultrasonics **40**, 873–878 (2002).
[45]F. A. Duck, *Physical Properties of Tissue* (Academic, San Diego, CA, 1990), Chap. 4.
[46]S. M. van der Meer, B. Dollet, M. M. Voormolen, C. T. Chin, A. Bouakaz, N. de Jong, M. Versluis, and D. Lohse, "Microbubble spectroscopy of ultrasound contrast agents," J. Acoust. Soc. Am. **121**, 648–656 (2007).
[47]J. A. Brown, F. S. Foster, A. Needles, E. Chérin, and G. R. Lockwood, "Fabrication and performance of a 40-MHz linear array based on a 1–3 composite with geometric elevation focusing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1888–1894 (2007).
[48]J. Naze Tjøtta, S. Tjøtta, and E. H. Vefring, "Effects of focusing on the nonlinear interaction between two collinear finite amplitude sound beams," J. Acoust. Soc. Am. **89**, 1017–1027 (1991).
[49]A. Needles, D. E. Goertz, R. Karshafian, E. Chérin, A. S. Brown, P. N. Burns, and F. S. Foster, "High-frequency subharmonic pulsed-wave Doppler and color flow imaging of microbubble contrast agents," Ultrasound Med. Biol. **34**, 1139–1151 (2008).
[50]M. Emmer, A. van Wamel, D. E. Goertz, and N. de Jong, "The onset of microbubble vibration," Ultrasound Med. Biol. **33**, 941–949 (2007).
[51]Y.-S. Lee, "Numerical solution of the KZK equation for pulsed finite amplitude sound beams in thermoviscous fluids," Ph.D. thesis, University of Texas, Austin, TX (1993).

Acoustic microstreaming around an encapsulated particle

Alexander A. Doinikov[a] and Ayache Bouakaz
INSERM U930 CNRS ERL 3106, Université François Rabelais, CHU Bretonneau, 2 Boulevard Tonnellé, 37044 Tours Cedex 9, France

(Received 10 December 2009; accepted 15 December 2009)

A theory is suggested to calculate acoustic microstreaming that develops around an encapsulated particle in an ultrasound field. It is assumed that the particle is suspended in an unbounded fluid and can have a gas core or a liquid core. Fluids outside and inside the particle are assumed to be viscous and compressible. No restrictions are imposed on the size of the particle relative to the sound and viscous wavelengths in the outer and inner fluids, and all modes of the particle's motion, including the surface ones, are taken into account. The shell of the particle is assumed to be an elastic viscous compressible medium. Solutions for the sound field within the shell are also valid for any ratio between the particle size and the sound and viscous wavelengths in the shell material. Numerical examples for the cases of an encapsulated gas bubble and an encapsulated particle with a liquid core are presented. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290997]

PACS number(s): 43.25.Nm, 43.25.Yw [RMW] Pages: 1218–1227

I. INTRODUCTION

Acoustic streaming produced by bubbles oscillating in a sound field has been a topic of research for many years (Nyborg, 1958, 1965; Elder, 1959; Davidson and Riley, 1971; Lee and Wang, 1990; Wu and Du, 1997; Longuet-Higgins, 1998). This attention is related to a diversity of technological and biomedical applications where this phenomenon plays a role. In recent years, in connection with clinical applications of contrast agent microbubbles, interest to acoustic microstreaming generated by encapsulated bubbles has arisen. Liu and Wu (2009) developed a theory to calculate the microstreaming velocity for an encapsulated bubble, taking account of two modes of the bubble's motion: a monopole (volume pulsation) and a dipole (translational harmonic vibration). Their theory also assumes that the wavelength of the incident sound field is much greater than the radius of the bubble, and the thickness of the viscous boundary layer surrounding the bubble is much smaller than the radius of the bubble. Liu and Wu (2009) used a theoretical approach that was suggested previously by Wu and Du (1997) for a free bubble. This approach is based on the assumption that viscous effects are essential only within a thin boundary layer while beyond this layer the fluid motion can be considered to be inviscid. However, Doinikov and Bouakaz (2010) showed that the approach suggested by Wu and Du (1997) may be applied only to very low-viscosity fluids, whose viscosity is far less than that of water. Under normal conditions, their solutions lead to a severe underestimation of the strength of acoustic streaming because dominant viscous terms in the first-order velocity field are determined incorrectly. The work of Liu and Wu (2009) encounters the same problem.

In this paper, a theory void of the restrictions of Liu and Wu's (2009) approach is developed. It is assumed that fluids outside and inside an encapsulated particle are viscous and compressible. Solutions which are derived for the scattered sound fields outside and inside the particle are valid for the entire volume of the outer and inner fluids. No restrictions are imposed on the size of the particle relative to the sound and viscous wavelengths in the outer and inner fluids, and all modes of the particle's motion, including the surface ones, are taken into account. The shell of the particle is assumed to be an elastic viscous compressible medium. Solutions for the sound field within the shell are also valid for any ratio between the particle size and the sound and viscous wavelengths in the shell material.

[a]Author to whom correspondence should be addressed. Electronic mail: doinikov@bsu.by

II. THEORY

In the process of the derivation of acoustic radiation force acting on a spherical particle of arbitrary internal structure in a sound wave field, Doinikov (1994, 1997) obtained solutions to the equations of acoustic streaming developing around the particle. These solutions impose no restrictions on the size of the particle with respect to the sound and viscous wavelengths in the ambient fluid and take into account all modes of the particle's motion: volume pulsation, translation, and shape oscillations. The velocity field of acoustic streaming is expressed in terms of linear scattering coefficients which are determined by the particle's internal structure. To apply Doinikov's (1994, 1997) theory to an encapsulated particle, linear scattering coefficients for this particle should be calculated. In Secs. II A–II E, we carry out this calculation. Section II F provides the equations of acoustic streaming in which the calculated scattering coefficients are used.

A. Acoustic field in the ambient liquid

The first-order velocity in the ambient liquid is represented as

$$\boldsymbol{v}_l = \nabla\varphi_I + \nabla\varphi_l + \nabla\times\boldsymbol{\psi}_l, \tag{1}$$

where φ_I is the scalar velocity potential of the incident acoustic field, and φ_l and $\boldsymbol{\psi}_l$ are the scalar and the vorticity

 0001-4966/2010/127(3)/1218/10/$25.00

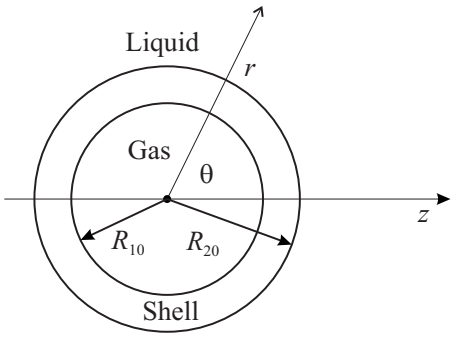


FIG. 1. Schematic sketch of an encapsulated particle.

velocity potentials of the scattered wave field. The potential φ_I is given by

$$\varphi_I = \exp(-i\omega t)\sum_{n=0}^{\infty} A_n j_n(k_l r) P_n(\cos\theta), \tag{2}$$

where j_n is the spherical Bessel function of order n, P_n is the Legendre polynomial of degree n, k_l is the wave number in the ambient liquid, and r and θ are the coordinates of the spherical coordinate system (r,θ,ε) whose origin is at the equilibrium center of the particle and the z axis (θ=0) lies in the direction of wave propagation, see Fig. 1. It should be emphasized that Eq. (2) describes an axially symmetric wave, and the particle is assumed to be located on the axis of the wave. The coefficients A_n specify the type of incident wave. For a plane traveling wave with the potential

$$\varphi_I = A\,\exp(i\boldsymbol{k}_l\cdot\boldsymbol{r} - i\omega t), \tag{3}$$

the coefficients A_n are given by

$$A_n = A(2n+1)i^n. \tag{4}$$

The wave number k_l is defined by

$$k_l = \frac{\omega}{c_l}\left[1 - \frac{i\omega}{\rho_l c_l^2}\left(\xi_l + \frac{4}{3}\eta_l\right)\right]^{-1/2}, \tag{5}$$

where c_l is the sound speed in the liquid, ρ_l is the equilibrium liquid density, η_l is the liquid shear viscosity, and ξ_l is the liquid bulk viscosity.

The velocity potentials of the scattered field are written as

$$\varphi_l = \exp(-i\omega t)\sum_{n=0}^{\infty} \alpha_n^{(l)} A_n h_n^{(1)}(k_l r) P_n(\cos\theta), \tag{6}$$

$$\boldsymbol{\psi}_l = \exp(-i\omega t)\boldsymbol{e}_\varepsilon \sum_{n=1}^{\infty} \beta_n^{(l)} A_n h_n^{(1)}(k_{lv} r) P_n^1(\cos\theta). \tag{7}$$

Here, $\alpha_n^{(l)}$ and $\beta_n^{(l)}$ are the linear scattering coefficients to be determined by the boundary conditions at the surface of the particle, $h_n^{(1)}$ is the spherical Hankel function of the first kind, $\boldsymbol{e}_\varepsilon$ is the unit azimuth vector of the spherical coordinate system introduced above, k_{lv} is the viscous wave number, and P_n^1 is the associated Legendre polynomial of the first order and degree n. Note that P_n^1 is defined here as

$$P_n^1(x) = (1-x^2)^{1/2} dP_n(x)/dx. \tag{8}$$

The viscous wave number k_{lv} is given by

$$k_{lv} = (1+i)/\delta_{lv}, \tag{9}$$

where $\delta_{lv} = \sqrt{2\nu_l/\omega}$ is the viscous penetration depth in the ambient liquid and $\nu_l = \eta_l/\rho_l$ is the liquid kinematic viscosity.

From Eqs. (1), (2), (6), and (7), the r and θ components of the liquid velocity are found to be

$$v_r^{(l)} = \exp(-i\omega t)\frac{1}{r}\sum_{n=0}^{\infty} A_n P_n(\cos\theta)[k_l r j_n'(k_l r) + \alpha_n^{(l)} k_l r h_n^{(1)\prime}(k_l r) + \beta_n^{(l)} n(n+1) h_n^{(1)}(k_{lv} r)], \tag{10}$$

$$v_\theta^{(l)} = -\exp(-i\omega t)\frac{1}{r}\sum_{n=1}^{\infty} A_n P_n^1(\cos\theta)[j_n(k_l r) + \alpha_n^{(l)} h_n^{(1)}(k_l r) + \beta_n^{(l)}(h_n^{(1)}(k_{lv} r) + k_{lv} r h_n^{(1)\prime}(k_{lv} r))], \tag{11}$$

where the prime denotes the differentiation with respect to the argument in parentheses.

The calculation of acoustic streaming also requires the expression for the first-order perturbation of the liquid density. Using the continuity equation, the liquid density perturbation can be written as

$$\tilde{\rho}_l = -\frac{i\rho_l}{\omega}\boldsymbol{\nabla}\cdot\boldsymbol{v}_l = \exp(-i\omega t)\frac{i\rho_l k_l^2}{\omega}\sum_{n=0}^{\infty} A_n P_n(\cos\theta) \times[j_n(k_l r) + \alpha_n^{(l)} h_n^{(1)}(k_l r)]. \tag{12}$$

If the particle is absent, the liquid density perturbation caused by the incident field is given by

$$\tilde{\rho}_I = i\rho_l k_l^2 \varphi_I/\omega. \tag{13}$$

B. Acoustic field in the encapsulating shell

The motion of the shell is described by the following equation (Landau and Lifshitz, 1986):

$$\rho_s \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ij}^{(s)}}{\partial x_j}, \tag{14}$$

where ρ_s is the equilibrium density of the shell, $\boldsymbol{u}$ is the displacement vector, $\sigma_{ij}^{(s)}$ is the shell stress tensor, and summation over double indices is implied. Assuming that the shell material possesses elasticity and viscosity, the shell stress tensor can be written as $\sigma_{ij}^{(s)} = \sigma_{ij}^{(\mathrm{el})} + \sigma_{ij}^{(\mathrm{vis})}$, where the elastic and the viscous parts are specified by Hooke's elastic law and Newton's viscous law,

$$\sigma_{ij}^{(\mathrm{el})} = \mu_s\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \left(K_s - \frac{2}{3}\mu_s\right)\delta_{ij}\frac{\partial u_k}{\partial x_k}, \tag{15}$$

$$\sigma_{ij}^{(\mathrm{vis})} = \eta_s\left(\frac{\partial v_i^{(s)}}{\partial x_j} + \frac{\partial v_j^{(s)}}{\partial x_i}\right) + \left(\xi_s - \frac{2}{3}\eta_s\right)\delta_{ij}\frac{\partial v_k^{(s)}}{\partial x_k}. \tag{16}$$

Here μ_s is the shell shear modulus, K_s is the shell bulk modulus, δ_{ij} is the Kronecker delta, η_s is the shell shear viscosity, ξ_s is the shell bulk viscosity, and $\boldsymbol{v}^{(s)}$ is the velocity within the shell material.

Substituting Eqs. (15) and (16) into Eq. (14) and taking into account that $\boldsymbol{v}^{(s)}=\partial \boldsymbol{u}/\partial t$ and the incident field depends on time as $\exp(-i\omega t)$, one obtains

$$\rho_s\omega^2\boldsymbol{u} + (\mu_s - i\omega\eta_s)\Delta\boldsymbol{u} + [K_s + \mu_s/3 - i\omega(\xi_s + \eta_s/3)]\nabla(\nabla\cdot\boldsymbol{u}) = 0. \tag{17}$$

The displacement vector can be represented as

$$\boldsymbol{u} = \nabla\varphi_s + \nabla\times\boldsymbol{\psi}_s. \tag{18}$$

Substitution of Eq. (18) into Eq. (17) yields

$$\Delta\varphi_s + k_s^2\varphi_s = 0, \tag{19}$$

$$\Delta\boldsymbol{\psi}_s + k_{sv}^2\boldsymbol{\psi}_s = 0, \tag{20}$$

where the wave numbers are given by

$$k_s = \left(\frac{\rho_s\omega^2}{K_s + 4\mu_s/3 - i\omega(\xi_s + 4\eta_s/3)}\right)^{1/2}, \tag{21}$$

$$k_{sv} = \left(\frac{\rho_s\omega^2}{\mu_s - i\omega\eta_s}\right)^{1/2}. \tag{22}$$

The axisymmetric solutions to Eqs. (19) and (20) are written as

$$\varphi_s = \exp(-i\omega t)\sum_{n=0}^{\infty} A_n[\alpha_{1n}^{(s)}h_n^{(1)}(k_s r) + \alpha_{2n}^{(s)}h_n^{(2)}(k_s r)]P_n(\cos\theta), \tag{23}$$

$$\boldsymbol{\psi}_s = \exp(-i\omega t)\boldsymbol{e}_\varepsilon\sum_{n=1}^{\infty} A_n[\beta_{1n}^{(s)}h_n^{(1)}(k_{sv}r) + \beta_{2n}^{(s)}h_n^{(2)}(k_{sv}r)] \times P_n^1(\cos\theta), \tag{24}$$

where $h_n^{(2)}$ is the spherical Hankel function of the second kind, and the coefficients $\alpha_{1n}^{(s)}$, $\alpha_{2n}^{(s)}$, $\beta_{1n}^{(s)}$, and $\beta_{2n}^{(s)}$ are determined by the boundary conditions at the outer and the inner surfaces of the shell.

Substituting Eqs. (23) and (24) into Eq. (18), one finds the r and θ components of the displacement vector to be

$$u_r = \exp(-i\omega t)\frac{1}{r}\sum_{n=0}^{\infty} A_n P_n(\cos\theta)\{\alpha_{1n}^{(s)}k_s r h_n^{(1)\prime}(k_s r) + \alpha_{2n}^{(s)}k_s r h_n^{(2)\prime}(k_s r) + n(n+1)[\beta_{1n}^{(s)}h_n^{(1)}(k_{sv}r) + \beta_{2n}^{(s)}h_n^{(2)}(k_{sv}r)]\}, \tag{25}$$

$$u_\theta = -\exp(-i\omega t)\frac{1}{r}\sum_{n=1}^{\infty} A_n P_n^1(\cos\theta)\{\alpha_{1n}^{(s)}h_n^{(1)}(k_s r) + \alpha_{2n}^{(s)}h_n^{(2)} \times (k_s r) + \beta_{1n}^{(s)}[h_n^{(1)}(k_{sv}r) + k_{sv}r h_n^{(1)\prime}(k_{sv}r)] + \beta_{2n}^{(s)}[h_n^{(2)}(k_{sv}r) + k_{sv}r h_n^{(2)\prime}(k_{sv}r)]\}. \tag{26}$$

C. Acoustic field in the inner medium

The inner medium of the particle is assumed to be a compressible viscous fluid. That can be a gas or a liquid. Although all equations obtained below can be also applied to a particle with a liquid core, the inner fluid will be referred to as a gas hereafter since the case of an encapsulated gas bubble is of primary interest to us.

The first-order velocity in the inner medium (gas) is represented as

$$\boldsymbol{v}_g = \nabla\varphi_g + \nabla\times\boldsymbol{\psi}_g. \tag{27}$$

For the velocity potentials, which should be finite at $r=0$, one has

$$\varphi_g = \exp(-i\omega t)\sum_{n=0}^{\infty}\alpha_n^{(g)}A_n j_n(k_g r)P_n(\cos\theta), \tag{28}$$

$$\boldsymbol{\psi}_g = \exp(-i\omega t)\boldsymbol{e}_\varepsilon\sum_{n=1}^{\infty}\beta_n^{(g)}A_n j_n(k_{gv}r)P_n^1(\cos\theta), \tag{29}$$

where the coefficients $\alpha_n^{(g)}$ and $\beta_n^{(g)}$ are determined by the boundary conditions at the inner surface of the shell, and the wave numbers are given by

$$k_g = \frac{\omega}{c_g}\left[1 - \frac{i\omega}{\rho_g c_g^2}\left(\xi_g + \frac{4}{3}\eta_g\right)\right]^{-1/2}, \tag{30}$$

$$k_{gv} = (1+i)/\delta_{gv}. \tag{31}$$

Here c_g is the sound speed in the gas, ρ_g is the equilibrium gas density, η_g is the gas shear viscosity, ξ_g is the gas bulk viscosity, $\delta_{gv}=\sqrt{2\nu_g/\omega}$ is the viscous penetration depth in the gas, and $\nu_g=\eta_g/\rho_g$ is the gas kinematic viscosity.

Substituting Eqs. (28) and (29) into Eq. (27), one finds the r and θ components of the gas velocity to be

$$v_r^{(g)} = \exp(-i\omega t)\frac{1}{r}\sum_{n=0}^{\infty}A_n P_n(\cos\theta)[\alpha_n^{(g)}k_g r j_n'(k_g r) + \beta_n^{(g)}n(n+1)j_n(k_{gv}r)], \tag{32}$$

$$v_\theta^{(g)} = -\exp(-i\omega t)\frac{1}{r}\sum_{n=1}^{\infty}A_n P_n^1(\cos\theta)[\alpha_n^{(g)}j_n(k_g r) + \beta_n^{(g)}(j_n(k_{gv}r) + k_{gv}r j_n'(k_{gv}r))]. \tag{33}$$

D. Boundary conditions

The boundary conditions at the outer and the inner surfaces of the shell are written as

$$v_r^{(l)} = v_r^{(s)}, \quad v_\theta^{(l)} = v_\theta^{(s)} \quad \text{at} \quad r = R_{20}, \tag{34}$$

$$v_r^{(s)} = v_r^{(g)}, \quad v_\theta^{(s)} = v_\theta^{(g)} \quad \text{at} \quad r = R_{10}, \tag{35}$$

$$\sigma_{rr}^{(l)} - \bar{p}_{\text{st}}^{(l)} = \sigma_{rr}^{(s)}, \quad \sigma_{r\theta}^{(l)} = \sigma_{r\theta}^{(s)} \quad \text{at} \quad r = R_{20}, \tag{36}$$

$$\sigma_{rr}^{(s)} = \sigma_{rr}^{(g)} - \tilde{p}_{\mathrm{st}}^{(g)}, \quad \sigma_{r\theta}^{(s)} = \sigma_{r\theta}^{(g)} \quad \text{at } r = R_{10}. \tag{37}$$

Here R_{10} and R_{20} are, respectively, the inner and the outer radii of the shell at rest, σ_{rr} and $\sigma_{r\theta}$ are the normal and the tangential components of the respective stress tensors, and $\tilde{p}_{\mathrm{st}}^{(l)}$ and $\tilde{p}_{\mathrm{st}}^{(g)}$ are the surface tension pressures at the liquid-shell and the shell-gas interfaces, respectively.

The r and θ components of the velocities are given by the equations derived in Secs. II A–II C. The quantity $\sigma_{rr}^{(l)}$ is calculated by (Landau and Lifshitz, 1987)

$$\sigma_{rr}^{(l)} = -p_l + 2\eta_l \frac{\partial v_r^{(l)}}{\partial r} + \left(\xi_l - \frac{2}{3}\eta_l\right) \boldsymbol{\nabla} \cdot \boldsymbol{v}_l, \tag{38}$$

where p_l is the first-order perturbed pressure in the ambient liquid. Using the continuity equation and the fact that $\boldsymbol{\nabla} \cdot \boldsymbol{v}_l = -k_l^2(\varphi_I + \varphi_l)$, p_l can be written as

$$p_l = -ic_l^2 \rho_l \boldsymbol{\nabla} \cdot \boldsymbol{v}_l / \omega = ik_l^2 c_l^2 \rho_l (\varphi_I + \varphi_l)/\omega. \tag{39}$$

Accordingly, Eq. (38) takes the final form

$$\sigma_{rr}^{(l)} = 2\eta_l \frac{\partial v_r^{(l)}}{\partial r} - \left(\frac{ic_l^2 \rho_l}{\omega} + \xi_l - \frac{2}{3}\eta_l\right) k_l^2 (\varphi_I + \varphi_l). \tag{40}$$

To calculate $\tilde{p}_{\mathrm{st}}^{(l)}$, the perturbed outer surface of the particle is represented as

$$R_2(\theta, t) = R_{20} + \exp(-i\omega t) \sum_{n=0}^{\infty} s_n^{(l)} P_n(\cos\theta). \tag{41}$$

Then the surface tension pressure on this surface can be written as (Jackson *et al.*, 1988)

$$p_{\mathrm{st}}^{(l)} = \frac{2\sigma_l}{R_{20}} + \exp(-i\omega t) \frac{\sigma_l}{R_{20}^2} \sum_{n=0}^{\infty} (n-1)(n+2) s_n^{(l)} P_n(\cos\theta), \tag{42}$$

where σ_l is the surface tension coefficient at the liquid-shell interface. The coefficients $s_n^{(l)}$ are found by using the following equality:

$$v_r^{(l)}(r = R_{20}) = \frac{\partial R_2(\theta, t)}{\partial t} = -i\omega \exp(-i\omega t) \sum_{n=0}^{\infty} s_n^{(l)} P_n(\cos\theta). \tag{43}$$

Substituting Eq. (10) into Eq. (43), one obtains

$$s_n^{(l)} = \frac{iA_n}{\omega R_{20}} [k_l R_{20} j_n'(k_l R_{20}) + \alpha_n^{(l)} k_l R_{20} h_n^{(1)\prime}(k_l R_{20}) + \beta_n^{(l)} n(n+1) h_n^{(1)}(k_{lv} R_{20})]. \tag{44}$$

In Eq. (36), the time-dependent part of $p_{\mathrm{st}}^{(l)}$ is required, which is indicated by the tilde. This part is given by

$$\tilde{p}_{\mathrm{st}}^{(l)} = \exp(-i\omega t) \frac{\sigma_l}{R_{20}^2} \sum_{n=0}^{\infty} (n-1)(n+2) s_n^{(l)} P_n(\cos\theta). \tag{45}$$

The quantity $\sigma_{r\theta}^{(l)}$ is calculated by (Landau and Lifshitz, 1987)

$$\sigma_{r\theta}^{(l)} = \eta_l \left(\frac{1}{r}\frac{\partial v_r^{(l)}}{\partial \theta} + \frac{\partial v_\theta^{(l)}}{\partial r} - \frac{v_\theta^{(l)}}{r}\right). \tag{46}$$

The expressions for the shell stress components are found from Eqs. (15) and (16) to be

$$\sigma_{rr}^{(s)} = 2(\mu_s - i\omega\eta_s)\frac{\partial u_r}{\partial r} - \left[K_s - \frac{2}{3}\mu_s - i\omega\left(\xi_s - \frac{2}{3}\eta_s\right)\right] k_s^2 \varphi_s, \tag{47}$$

$$\sigma_{r\theta}^{(s)} = (\mu_s - i\omega\eta_s)\left(\frac{1}{r}\frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right). \tag{48}$$

By analogy with those for the ambient liquid, the expressions for the stress components in the gas can be written as

$$\sigma_{rr}^{(g)} = 2\eta_g \frac{\partial v_r^{(g)}}{\partial r} - \left(\frac{ic_g^2 \rho_g}{\omega} + \xi_g - \frac{2}{3}\eta_g\right) k_g^2 \varphi_g, \tag{49}$$

$$\tilde{p}_{\mathrm{st}}^{(g)} = \exp(-i\omega t) \frac{\sigma_g}{R_{10}^2} \sum_{n=0}^{\infty} (n-1)(n+2) s_n^{(g)} P_n(\cos\theta), \tag{50}$$

$$s_n^{(g)} = \frac{iA_n}{\omega R_{10}} [\alpha_n^{(g)} k_g R_{10} j_n'(k_g R_{10}) + \beta_n^{(g)} n(n+1) j_n(k_{gv} R_{10})], \tag{51}$$

$$\sigma_{r\theta}^{(g)} = \eta_g \left(\frac{1}{r}\frac{\partial v_r^{(g)}}{\partial \theta} + \frac{\partial v_\theta^{(g)}}{\partial r} - \frac{v_\theta^{(g)}}{r}\right). \tag{52}$$

Substituting the expressions for the velocity and stress components into the boundary conditions, one can calculate the unknown scattering coefficients.

E. Calculation of the scattering coefficients

The boundary conditions, Eqs. (34)–(37), give a system of eight algebraic equations in the unknowns $\alpha_n^{(l)}$, $\beta_n^{(l)}$, $\alpha_{1n}^{(s)}$, $\alpha_{2n}^{(s)}$, $\beta_{1n}^{(s)}$, $\beta_{2n}^{(s)}$, $\alpha_n^{(g)}$, and $\beta_n^{(g)}$. The system can be written in matrix form as

$$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}, \tag{53}$$

where the elements of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are given in the Appendix, and the matrix $\boldsymbol{X}$ consists of the scattering coefficients which are arranged in the same order as indicated above. Note that the coefficients $\beta_0^{(l)}$, $\beta_{10}^{(s)}$, $\beta_{20}^{(s)}$, and $\beta_0^{(g)}$ are identically zero. Therefore for $n=0$ system (53) reduces to four equations in the unknowns $\alpha_0^{(l)}$, $\alpha_{10}^{(s)}$, $\alpha_{20}^{(s)}$, and $\alpha_0^{(g)}$,

$$\boldsymbol{A}_0 \cdot \boldsymbol{X}_0 = \boldsymbol{B}_0. \tag{54}$$

The elements of the matrices $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are also given in the Appendix, and the matrix $\boldsymbol{X}_0$ is formed by the four coefficients pointed out above.

For the calculation of acoustic streaming, only $\alpha_n^{(l)}$ and $\beta_n^{(l)}$ are needed. They are calculated by

$$\alpha_0^{(l)} = \det(\boldsymbol{A}_0^{(1)})/\det(\boldsymbol{A}_0), \quad \alpha_n^{(l)} = \det(\boldsymbol{A}^{(1)})/\det(\boldsymbol{A}),$$

$$\beta_n^{(l)} = \det(\boldsymbol{A}^{(2)})/\det(\boldsymbol{A}), \tag{55}$$

where $\boldsymbol{A}^{(m)}$ denotes the matrix $\boldsymbol{A}$ in which the mth column is replaced with the elements of the matrix $\boldsymbol{B}$.

F. Acoustic streaming

According to the results of Doinikov (1994, 1997), the velocity of acoustic streaming can be written as

$$\boldsymbol{V} = \boldsymbol{\nabla}\Phi + \boldsymbol{\nabla} \times \boldsymbol{\Psi}, \tag{56}$$

where the scalar and vorticity potentials are given by

$$\Phi = \sum_{n=0}^{\infty} \Phi_n(r/R_{20}) P_n(\cos\,\theta), \tag{57}$$

$$\boldsymbol{\Psi} = \boldsymbol{e}_{\varepsilon} \sum_{n=1}^{\infty} \Psi_n(r/R_{20}) P_n^1(\cos\,\theta). \tag{58}$$

Substituting Eqs. (57) and (58) into Eq. (56), one finds the r and θ components of the acoustic streaming velocity to be

$$V_r = \frac{1}{R_{20}} \sum_{n=0}^{\infty} \left[\Phi_n'(x) + \frac{n(n+1)}{x} \Psi_n(x) \right] P_n(\cos\,\theta), \tag{59}$$

$$V_\theta = -\frac{1}{R_{20}} \sum_{n=1}^{\infty} \left[\frac{1}{x}\Phi_n(x) + \frac{1}{x}\Psi_n(x) + \Psi_n'(x) \right] P_n^1(\cos\,\theta), \tag{60}$$

where $x = r/R_{20}$ and the prime denotes the differentiation with respect to the argument in parentheses.

The functions $\Phi_n(x)$ and $\Psi_n(x)$ are calculated by

$$\Phi_n(x) = \frac{1}{x^{n+1}}[K_n(x,n+2) - C_{\varphi n}] - x^n[K_n(x,1-n) - K_n(\infty,1-n)], \tag{61}$$

$$\Psi_n(x) = \frac{1}{2(2n+3)} \left\{ \frac{1}{x^{n+1}} N_n(x,n+3) - x^{n+2}[N_n(x,-n) - N_n(\infty,-n)] \right\} + \frac{1}{2(2n-1)} \{x^n[N_n(x,2-n) - N_n(\infty,2-n)] - x^{1-n}[N_n(x,n+1) - C_{\psi n}]\}, \tag{62}$$

where

$$K_n(x,m) = \int_1^x y^m \mu_n(y)\,dy, \tag{63}$$

$$\mu_n(r/R_{20}) = \frac{R_{20}^2}{4\rho_l} \int_0^{\pi} \mathrm{Re}\{\boldsymbol{\nabla} \cdot (\tilde{\rho}_l^* \boldsymbol{v}_l - \tilde{\rho}_I^* \boldsymbol{v}_I)\} P_n(\cos\,\theta) \sin\,\theta\,d\theta, \tag{64}$$

$$N_n(x,m) = \int_1^x y^m [\chi_{rn}(y) + m\chi_{\theta n}(y)]\,dy, \tag{65}$$

$$\chi_{rn}(r/R_{20}) = \int_0^{\pi} W_r(r/R_{20},\theta) P_n(\cos\,\theta) \sin\,\theta\,d\theta, \tag{66}$$

$$\chi_{\theta n}(r/R_{20}) = -\frac{1}{n(n+1)} \int_0^{\pi} W_\theta(r/R_{20},\theta) P_n^1(\cos\,\theta) \sin\,\theta\,d\theta, \tag{67}$$

$$W(r/R_{20},\theta) = \frac{R_{20}^3}{4\nu_l} \mathrm{Re}\{\boldsymbol{v}_l^*(\boldsymbol{\nabla}\cdot\boldsymbol{v}_l) + (\boldsymbol{v}_l^* \cdot \boldsymbol{\nabla})\boldsymbol{v}_l - \boldsymbol{v}_I^*(\boldsymbol{\nabla}\cdot\boldsymbol{v}_I) - (\boldsymbol{v}_I^* \cdot \boldsymbol{\nabla})\boldsymbol{v}_I\}, \tag{68}$$

$$C_{\varphi n} = \frac{n(2n-1)}{2(n+1)} K_n(\infty,1-n) + \frac{n(2n+1)}{4(2n+3)} N_n(\infty,-n) - \frac{n}{4} N_n(\infty,2-n) + \frac{(2n+1)(n-2)}{2(n+1)} a_n + \frac{n(2n+1)}{2} b_n, \tag{69}$$

$$C_{\psi n} = (4n^2-1) \left[\frac{N_n(\infty,2-n)}{2(2n-1)} - \frac{N_n(\infty,-n)}{2(2n+1)} - \frac{K_n(\infty,1-n) + a_n}{n+1} - b_n \right], \tag{70}$$

$$a_n = \int_0^{\pi} U_r(R_{20},\theta) P_n(\cos\,\theta) \sin\,\theta\,d\theta, \tag{71}$$

$$b_n = -\frac{1}{n(n+1)} \int_0^{\pi} U_\theta(R_{20},\theta) P_n^1(\cos\,\theta) \sin\,\theta\,d\theta, \tag{72}$$

$$\boldsymbol{U}(r,\theta) = \frac{R_{20}}{2}\boldsymbol{V}_I - \frac{R_{20}}{4\omega^2} \mathrm{Re}\left\{ \left(\frac{\partial v_r^{(l)}}{\partial t} \right)^* \frac{\partial \boldsymbol{v}_l}{\partial r} \right\}. \tag{73}$$

Here, the asterisk denotes the complex conjugate, Re means "the real part of," and $\boldsymbol{V}_I$ is the velocity produced by the incident sound field in the ambient liquid when the particle is absent. This velocity results from the space attenuation of the incident sound wave. For a plane traveling wave, $\boldsymbol{V}_I$ is given by

$$\boldsymbol{V}_I = -\frac{|k_l A|^2}{2\omega} \mathrm{Re}\{\boldsymbol{k}_l\} \exp(i\boldsymbol{k}_l \cdot \boldsymbol{r} - i\boldsymbol{k}_l^* \cdot \boldsymbol{r}). \tag{74}$$

According to Eq. (60) in Doinikov (1997), the radial stress produced by acoustic streaming can be calculated by

$$S_r = -\frac{1}{2\rho_l} \left(\xi_l + \frac{4\eta_l}{3} \right) \mathrm{Re}\{\boldsymbol{\nabla} \cdot (\tilde{\rho}_l^* \boldsymbol{v}_l - \tilde{\rho}_I^* \boldsymbol{v}_I)\} - \frac{\eta_l}{R_{20}^2} \sum_{n=0}^{\infty} Q_{rn}(r/R_{20}) P_n(\cos\,\theta), \tag{75}$$

where

$$Q_{rn}(x) = \frac{n}{x^{n+1}}[N_n(x,n+1) - C_{\psi n}] + (n+1)x^n[N_n(x,-n) - N_n(\infty,-n)]. \tag{76}$$

Substituting Eqs. (59) and (60) into the formula

$$S_\theta = \eta_l \left(\frac{1}{r} \frac{\partial V_r}{\partial \theta} + \frac{\partial V_\theta}{\partial r} - \frac{V_\theta}{r} \right), \tag{77}$$

one finds the tangential stress produced by acoustic streaming to be

$$S_\theta = \frac{\eta_l}{R_{20}^2} \sum_{n=1}^{\infty} Q_{\theta n}(r/R_{20}) P_n^1(\cos\theta), \tag{78}$$

where

$$Q_{\theta n}(x) = \frac{2}{x^2}\Phi_n(x) - \frac{2}{x}\Phi_n'(x) + \frac{2-2n(n+1)}{x^2}\Psi_n(x) + \frac{2}{x}\Psi_n'(x) + x^n[N_n(x,-n) - N_n(\infty,-n)] - \frac{1}{x^{n+1}}[N_n(x,n+1) - C_{\psi m}]. \tag{79}$$

It should be noted that although the equations for acoustic streaming look cumbersome, their numerical calculation is straightforward.

III. NUMERICAL EXAMPLES

As a first example, let us consider an encapsulated gas bubble with $R_{20}=2$ μm and the shell thickness $R_s=10$ nm driven at $f=1$ MHz. The values of the other parameters used in this calculation are the following: $\rho_l=1000$ kg/m^3, $\eta_l=0.001$ Pa s, $\xi_l=0.0026$ Pa s, $\sigma_l=0.005$ N/m, $c_l=1500$ m/s, $\rho_s=1100$ kg/m^3, $\mu_s=15$ MPa, $\eta_s=0.05$ Pa s, $\xi_s=0$, $\eta_g=1.8\times10^{-5}$ Pa s, $\xi_g=1.6\times10^{-5}$ Pa s, and $\sigma_g=0.04$ N/m. These data correspond to the case considered by Liu and Wu (2009). The values of the shell parameters were adopted by Liu and Wu (2009) from Church's (1995) work. The equilibrium gas pressure inside the bubble is calculated by $P_g=P_0+2\sigma_l/R_{20}+2\sigma_g/R_{10}$, where $P_0=101.3$ kPa is the hydrostatic pressure in the liquid. The equilibrium gas density is then calculated as $\rho_g=\rho_{gA}P_g/P_A$, where $\rho_{gA}=1.2$ kg/m^3 is the gas density at the atmospheric pressure $P_A=101.3$ kPa. Finally, the sound speed in the gas is defined by the equation $c_g=\sqrt{\gamma_g P_g/\rho_g}$ which results from the adiabatic law, γ_g denoting the special heat ratio. Liu and Wu (2009) did not indicate the value of the acoustic pressure amplitude which was used in their simulation. Supposedly, they used 100 kPa. In our simulations, the same value is used. To make a comparison with Liu and Wu's (2009) results, one more point should be taken into consideration. Our theory allows the shell material to be compressible, which means that the shell bulk modulus K_s is finite. Liu and Wu (2009) assumed the shell material to be incompressible, which means that $K_s\to\infty$. To simulate this situation, however, it is sufficient to set K_s to be much greater than the shell shear modulus μ_s. We have used $K_s=10^{10}$ Pa.

Figure 2 shows the radial and the tangential acoustic streaming velocities versus θ, and Fig. 3 provides plots of the radial and tangential stresses produced by the acoustic streaming. The values of the velocities and the stresses are calculated at $r=R_{20}+0.5\delta_{lv}$ as in Liu and Wu's (2009) paper. In both figures, the solid lines correspond to an encapsulated bubble, and the dashed lines show the results for a free bubble of the same size. One can see that the acoustic streaming produced by the encapsulated bubble is weaker than that produced by the free bubble. Liu and Wu (2009) reported an opposite result. They stated that, in spite of the fact that the oscillation amplitude of the encapsulated bubble was smaller than that of the free bubble, the streaming velocity and the stress induced by the encapsulated bubble were considerably greater than those for the free bubble. They suggested that this enhancement is caused by the difference in the boundary conditions for the free and encapsulated bubbles. However, this argument does not look con-

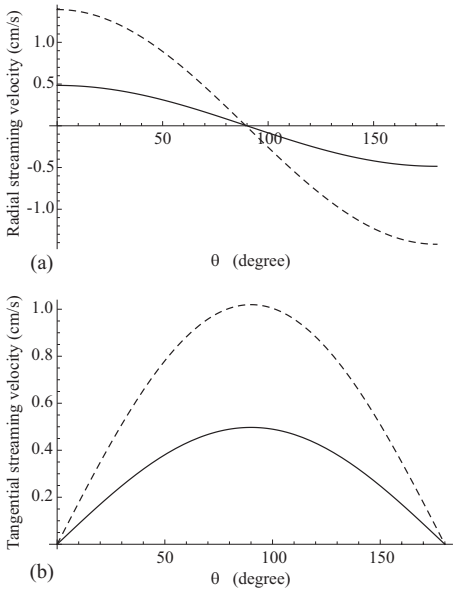


FIG. 2. Acoustic streaming velocity versus θ for an encapsulated gas bubble with $R_{20}=2$ μm and the shell thickness 10 nm. Excitation is a plane traveling wave with the frequency 1 MHz and the pressure amplitude 100 kPa. The dashed line corresponds to a free bubble of the same radius.

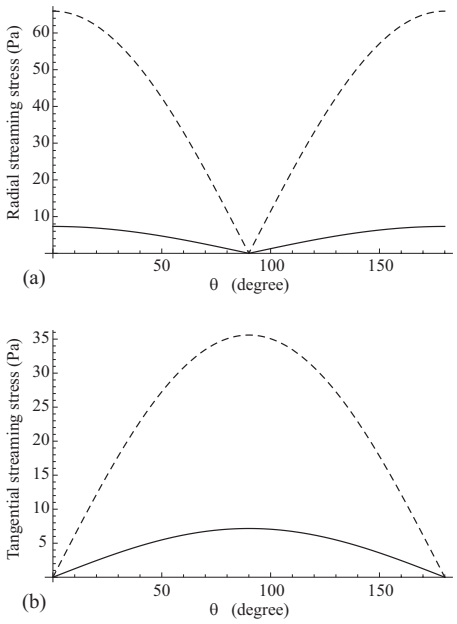


FIG. 3. Acoustic streaming stresses versus θ for the bubbles in Fig. 2. The dashed line corresponds to the free bubble.

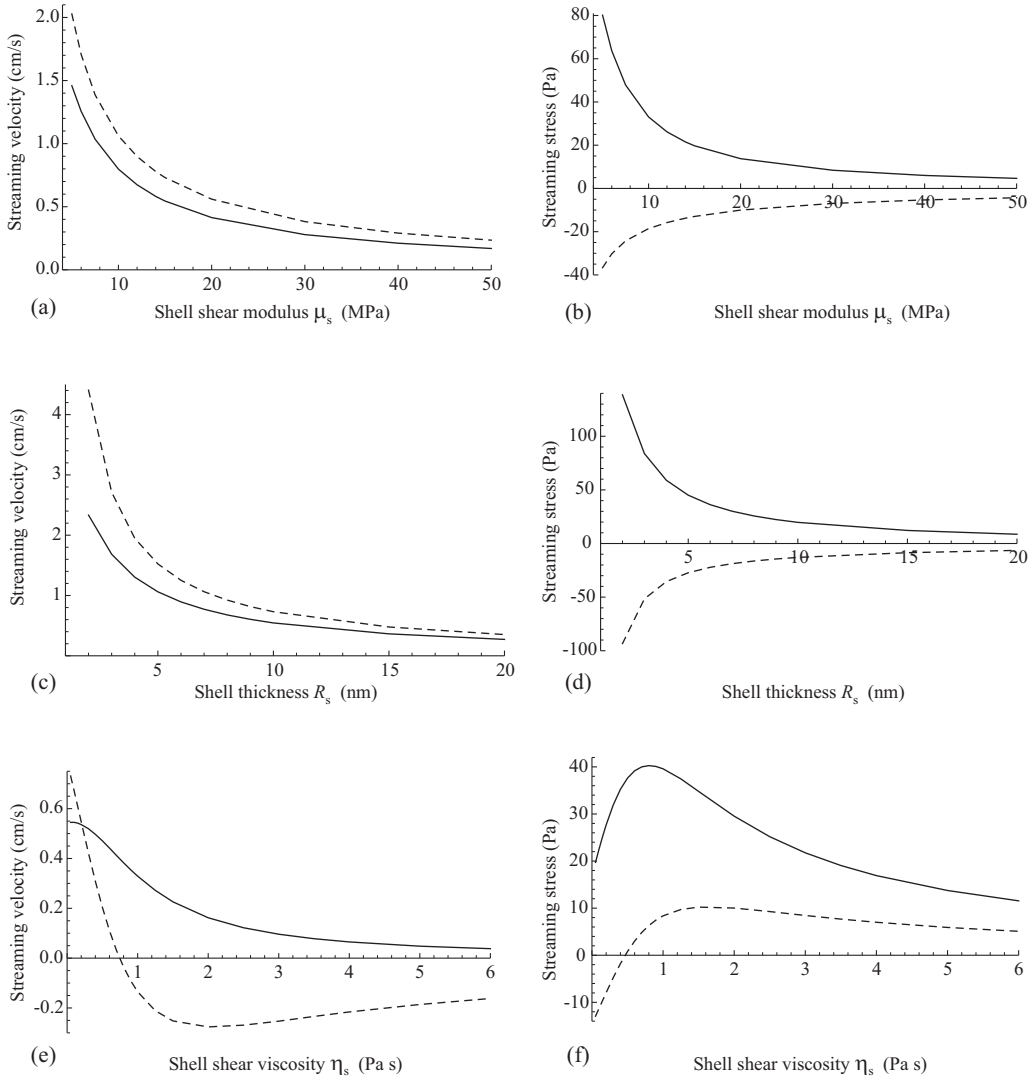


FIG. 4. Dependence of the maximum values of the streaming velocity and the streaming stress on the shell parameters. A bubble with $R_{20}=1.5$ μm is driven at 2.5 MHz and 100 kPa. The solid lines correspond to the radial velocity and stress components, and the dashed lines, to the tangential components. In (a) and (b), $R_s=10$ nm, $\eta_s=0.05$ Pa s; in (c) and (d), $\mu_s=15$ MPa, $\eta_s=0.05$ Pa s; and in (e) and (f), $\mu_s=15$ MPa, $R_s=10$ nm.

vincing as the increase in the acoustic streaming for the encapsulated bubble is extremely great. For example, according to Fig. 3(b) of Liu and Wu's (2009) paper, the radial streaming velocity for the encapsulated bubble is approximately 47 000 times greater than that for the free bubble, and Fig. 8(b) of their paper shows that the tangential stress for the encapsulated bubble is approximately 2400 times that for the free bubble. This erroneous finding is a consequence of the incorrect approximate approach which was used to calculate the first-order velocity field; see Doinikov and Bouakaz (2010). Figures 2 and 3 demonstrate that the correct calculation gives a quite expected result, namely, stronger oscillations produce a stronger acoustic streaming.

The effect of the shell parameters on acoustic streaming has been investigated. As could be expected, it is found that a stronger bubble oscillation leads to a stronger acoustic streaming. Thus a stronger acoustic streaming arises when a change in the shell parameters increases the bubble oscillation. For example, the oscillation increases when the resonance frequency of a bubble approaches the driving frequency. It is known that the resonance frequency increases with the shell shear modulus μ_s. Therefore, if a bubble is driven below resonance, a decrease in μ_s will enhance acoustic streaming and vice versa. This situation is illustrated in Figs. 4(a) and 4(b) which show the maximum values of the streaming velocity and the streaming stress as a function of the shell shear modulus μ_s. The simulations were carried out for a bubble with $R_{20}=1.5$ μm and $R_s=10$ nm driven at 2.5 MHz and 100 kPa. The other parameters are as indicated above. The values of the velocities and the stresses are calculated at $r=R_{20}+0.5\delta_{lv}$. The solid lines correspond to the radial components of the streaming velocity and the streaming stress, and the dashed lines, to the tangential components. Figures 4(c) and 4(d) show the effect of the shell thickness on acoustic streaming. It is known that the thicker the shell, the higher the resonance frequency. Therefore increasing the shell thickness gives the same effect as increasing the shell shear modulus. The effect of the shell viscosity is more complicated as the shell viscosity affects not only the resonance frequency of a bubble but also the damping and

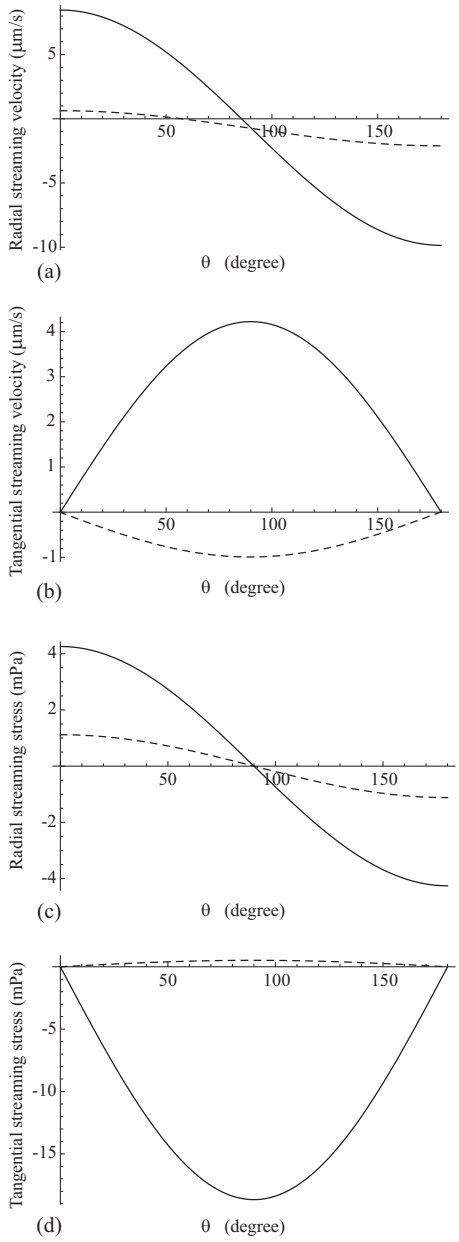


FIG. 5. Acoustic streaming velocities and stresses versus θ for an encapsulated particle with a liquid core. The dashed lines show the result of the calculation when the surface quadrupole mode is ignored.

the phase of the bubble oscillation. As a result, a change in the shell viscosity can even reverse the direction of acoustic streaming, as is seen in Fig. 4(e).

In the case of an encapsulated gas bubble, the radial oscillation mode is dominant. In combination with the translational mode, they make the main contribution to acoustic streaming. For example, for the bubble in Fig. 4, with the shell parameters $R_s=10$ nm, $\mu_s=15$ MPa, and $\eta_s=0.05$ Pa s, the ratio of the amplitude of the radial mode to R_{20} is 0.074, the same ratio for the translational mode is 0.0047, and that for the quadrupole mode is 3.4×10^{-5}. These figures show that the contribution of the surface modes is negligible. However, in the case of an encapsulated particle with a liquid core, the situation is different. As an example, let us consider a particle with $R_{20}=1.5$ μm and $R_s=10$ nm driven at 2.5 MHz and 100 kPa. The parameters of the ambient liquid and the shell are the same as indicated at the beginning of this section. Thus the only distinction from the bubble in Fig. 4 is that the particle under consideration has a liquid core. The data for the inner liquid are $\rho_g=880$ kg/m^3, $\eta_g=6.5\times10^{-4}$ Pa s, $\xi_g=0$, $\sigma_g=0.04$ N/m, $c_g=1321$ m/s, and $\gamma_g=1.1$. These parameters correspond to benzene. The subscript g is used here just as a sign of the inner medium. Under these conditions, the ratio of the amplitude of the radial mode to R_{20} is 2.1×10^{-5}, the ratio for the translational mode is 3.0×10^{-3}, for the quadrupole mode, 2.8×10^{-4}, and for the next surface mode, 4.5×10^{-8}. It follows that the main contribution to acoustic streaming comes from the interaction between the translational and the quadrupole modes. The velocity and stress fields of acoustic streaming that arises around the particle are shown in Fig. 5. The values of the velocities and the stresses are calculated at $r=R_{20}+0.5\delta_{lv}$. For comparison, the dashed lines show the result of calculation in the case that the quadrupole mode is dropped and only the radial and the translational modes are taken into account.

IV. CONCLUSIONS

A theory has been developed to calculate acoustic microstreaming that arises around an encapsulated particle in an ultrasound field. The particle was assumed to be suspended in an unbounded fluid. The surrounding fluid and a fluid inside the particle were assumed to be viscous and compressible. The shell of the particle was assumed to be an elastic viscous compressible medium. No restrictions were imposed on the size of the particle relative to the lengths of the sound and viscous waves that propagate in the surrounding fluid, the inner fluid, and the material of the shell. All modes of the particle's motion (volume pulsation, translation, and shape oscillations) were taken into account. The developed theory can be applied to an encapsulated gas bubble and an encapsulated particle with a liquid core. Numerical examples for both cases were presented.

ACKNOWLEDGMENT

Alexander Doinikov gratefully acknowledges the financial support from the le STUDIUM®, Institute for Advanced Studies (Orléans, France).

APPENDIX: ELEMENTS OF THE MATRICES *A* AND *B*

The elements of the matrix $\boldsymbol{A}_0$ $(n=0)$ are given by

$$a_{11}^{(0)}=-k_lR_{20}h_1^{(1)}(k_lR_{20}),\quad a_{12}^{(0)}=-i\omega k_sR_{20}h_1^{(1)}(k_sR_{20}),$$

$$a_{13}^{(0)}=-i\omega k_sR_{20}h_1^{(2)}(k_sR_{20}),\quad a_{14}^{(0)}=0,$$

$$a_{21}^{(0)}=0,\quad a_{22}^{(0)}=-i\omega k_sR_{10}h_1^{(1)}(k_sR_{10}),$$

$$a_{23}^{(0)}=-i\omega k_sR_{10}h_1^{(2)}(k_sR_{10}),\quad a_{24}^{(0)}=-k_gR_{10}j_1(k_gR_{10}),$$

$$a_{31}^{(0)} = -2\eta_l (k_l R_{20})^2 h_1^{(1)\prime}(k_l R_{20}) - \frac{2i\sigma_l}{\omega R_{20}} k_l R_{20} h_1^{(1)}(k_l R_{20}) - \left(\frac{ic_l^2 \rho_l}{\omega} + \xi_l - \frac{2}{3}\eta_l \right)(k_l R_{20})^2 h_0^{(1)}(k_l R_{20}),$$

$$a_{32}^{(0)} = 2(\mu_s - i\omega\eta_s)(k_s R_{20})^2 h_1^{(1)\prime}(k_s R_{20}) + \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{20})^2 h_0^{(1)}(k_s R_{20}),$$

$$a_{33}^{(0)} = 2(\mu_s - i\omega\eta_s)(k_s R_{20})^2 h_1^{(2)\prime}(k_s R_{20}) + \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{20})^2 h_0^{(2)}(k_s R_{20}),$$

$$a_{34}^{(0)} = 0, \quad a_{41}^{(0)} = 0,$$

$$a_{42}^{(0)} = 2(i\omega\eta_s - \mu_s)(k_s R_{10})^2 h_1^{(1)\prime}(k_s R_{10}) - \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{10})^2 h_0^{(1)}(k_s R_{10}),$$

$$a_{43}^{(0)} = 2(i\omega\eta_s - \mu_s)(k_s R_{10})^2 h_1^{(2)\prime}(k_s R_{10}) - \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{10})^2 h_0^{(2)}(k_s R_{10}),$$

$$a_{44}^{(0)} = 2\eta_g (k_g R_{10})^2 j_1'(k_g R_{10}) + \left(\frac{ic_g^2 \rho_g}{\omega} + \xi_g - \frac{2}{3}\eta_g \right) \times (k_g R_{10})^2 j_0(k_g R_{10}).$$

The elements of the matrix $\boldsymbol{A}$ $(n \geq 1)$ are given by

$$a_{11}^{(n)} = k_l R_{20} h_n^{(1)\prime}(k_l R_{20}), \quad a_{12}^{(n)} = n(n+1) h_n^{(1)}(k_{lv} R_{20}),$$

$$a_{13}^{(n)} = i\omega k_s R_{20} h_n^{(1)\prime}(k_s R_{20}), \quad a_{14}^{(n)} = i\omega k_s R_{20} h_n^{(2)\prime}(k_s R_{20}),$$

$$a_{15}^{(n)} = i\omega n(n+1) h_n^{(1)}(k_{sv} R_{20}),$$

$$a_{16}^{(n)} = i\omega n(n+1) h_n^{(2)}(k_{sv} R_{20}),$$

$$a_{17}^{(n)} = 0, \quad a_{18}^{(n)} = 0,$$

$$a_{21}^{(n)} = h_n^{(1)}(k_l R_{20}), \quad a_{22}^{(n)} = h_n^{(1)}(k_{lv} R_{20}) + k_{lv} R_{20} h_n^{(1)\prime}(k_{lv} R_{20}),$$

$$a_{23}^{(n)} = i\omega h_n^{(1)}(k_s R_{20}), \quad a_{24}^{(n)} = i\omega h_n^{(2)}(k_s R_{20}),$$

$$a_{25}^{(n)} = i\omega[h_n^{(1)}(k_{sv} R_{20}) + k_{sv} R_{20} h_n^{(1)\prime}(k_{sv} R_{20})],$$

$$a_{26}^{(n)} = i\omega[h_n^{(2)}(k_{sv} R_{20}) + k_{sv} R_{20} h_n^{(2)\prime}(k_{sv} R_{20})],$$

$$a_{27}^{(n)} = 0, \quad a_{28}^{(n)} = 0,$$

$$a_{31}^{(n)} = 0, \quad a_{32}^{(n)} = 0,$$

$$a_{33}^{(n)} = i\omega k_s R_{10} h_n^{(1)\prime}(k_s R_{10}), \quad a_{34}^{(n)} = i\omega k_s R_{10} h_n^{(2)\prime}(k_s R_{10}),$$

$$a_{35}^{(n)} = i\omega n(n+1) h_n^{(1)}(k_{sv} R_{10}),$$

$$a_{36}^{(n)} = i\omega n(n+1) h_n^{(2)}(k_{sv} R_{10}),$$

$$a_{37}^{(n)} = k_g R_{10} j_n'(k_g R_{10}), \quad a_{38}^{(n)} = n(n+1) j_n(k_{gv} R_{10}),$$

$$a_{41}^{(n)} = 0, \quad a_{42}^{(n)} = 0,$$

$$a_{43}^{(n)} = i\omega h_n^{(1)}(k_s R_{10}), \quad a_{44}^{(n)} = i\omega h_n^{(2)}(k_s R_{10}),$$

$$a_{45}^{(n)} = i\omega[h_n^{(1)}(k_{sv} R_{10}) + k_{sv} R_{10} h_n^{(1)\prime}(k_{sv} R_{10})], \quad a_{46}^{(n)} = i\omega[h_n^{(2)}(k_{sv} R_{10}) + k_{sv} R_{10} h_n^{(2)\prime}(k_{sv} R_{10})],$$

$$a_{47}^{(n)} = j_n(k_g R_{10}), \quad a_{48}^{(n)} = j_n(k_{gv} R_{10}) + k_{gv} R_{10} j_n'(k_{gv} R_{10}),$$

$$a_{51}^{(n)} = 2\eta_l (k_l R_{20})^2 h_n^{(1)\prime\prime}(k_l R_{20}) - \frac{i\sigma_l (n-1)(n+2)}{\omega R_{20}} k_l R_{20} h_n^{(1)\prime}(k_l R_{20}) - \left(\frac{ic_l^2 \rho_l}{\omega} + \xi_l - \frac{2}{3}\eta_l \right)(k_l R_{20})^2 h_n^{(1)}(k_l R_{20}),$$

$$a_{52}^{(n)} = 2\eta_l n(n+1)[k_{lv} R_{20} h_n^{(1)\prime}(k_{lv} R_{20}) - h_n^{(1)}(k_{lv} R_{20})] - \frac{i\sigma_l n(n^2-1)(n+2)}{\omega R_{20}} h_n^{(1)}(k_{lv} R_{20}),$$

$$a_{53}^{(n)} = 2(i\omega\eta_s - \mu_s)(k_s R_{20})^2 h_n^{(1)\prime\prime}(k_s R_{20}) + \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{20})^2 h_n^{(1)}(k_s R_{20}),$$

$$a_{54}^{(n)} = 2(i\omega\eta_s - \mu_s)(k_s R_{20})^2 h_n^{(2)\prime\prime}(k_s R_{20}) + \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_s R_{20})^2 h_n^{(2)}(k_s R_{20}),$$

$$a_{55}^{(n)} = 2(i\omega\eta_s - \mu_s) n(n+1)[k_{sv} R_{20} h_n^{(1)\prime}(k_{sv} R_{20}) - h_n^{(1)}(k_{sv} R_{20})],$$

$$a_{56}^{(n)} = 2(i\omega\eta_s - \mu_s) n(n+1)[k_{sv} R_{20} h_n^{(2)\prime}(k_{sv} R_{20}) - h_n^{(2)}(k_{sv} R_{20})],$$

$$a_{57}^{(n)} = 0, \quad a_{58}^{(n)} = 0,$$

$$a_{61}^{(n)} = 2\eta_l[k_l R_{20} h_n^{(1)\prime}(k_l R_{20}) - h_n^{(1)}(k_l R_{20})],$$

$$a_{62}^{(n)} = \eta_l\{(k_{lv} R_{20})^2 h_n^{(1)\prime\prime}(k_{lv} R_{20}) + [n(n+1) - 2] \times h_n^{(1)}(k_{lv} R_{20})\},$$

$$a_{63}^{(n)} = 2(i\omega\eta_s - \mu_s)[k_s R_{20} h_n^{(1)\prime}(k_s R_{20}) - h_n^{(1)}(k_s R_{20})],$$

$$a_{64}^{(n)} = 2(i\omega\eta_s - \mu_s)[k_s R_{20} h_n^{(2)\prime}(k_s R_{20}) - h_n^{(2)}(k_s R_{20})],$$

$$a_{65}^{(n)} = (i\omega\eta_s - \mu_s)\{(k_{sv} R_{20})^2 h_n^{(1)\prime\prime}(k_{sv} R_{20}) + [n(n+1) - 2] h_n^{(1)}(k_{sv} R_{20})\},$$

$$a_{66}^{(n)} = (i\omega\eta_s - \mu_s)\{(k_{sv} R_{20})^2 h_n^{(2)\prime\prime}(k_{sv} R_{20}) + [n(n+1) - 2] h_n^{(2)}(k_{sv} R_{20})\},$$

$$a_{67}^{(n)} = 0, \quad a_{68}^{(n)} = 0,$$

$$a_{71}^{(n)} = 0, \quad a_{72}^{(n)} = 0,$$

$$a_{73}^{(n)} = 2(\mu_s - i\omega\eta_s)(k_sR_{10})^2 h_n^{(1)\prime\prime}(k_sR_{10}) - \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_sR_{10})^2 h_n^{(1)}(k_sR_{10}),$$

$$a_{74}^{(n)} = 2(\mu_s - i\omega\eta_s)(k_sR_{10})^2 h_n^{(2)\prime\prime}(k_sR_{10}) - \left[K_s - \tfrac{2}{3}\mu_s - i\omega\left(\xi_s - \tfrac{2}{3}\eta_s\right)\right](k_sR_{10})^2 h_n^{(2)}(k_sR_{10}),$$

$$a_{75}^{(n)} = 2(\mu_s - i\omega\eta_s)n(n+1)[k_{sv}R_{10}h_n^{(1)\prime}(k_{sv}R_{10}) - h_n^{(1)}(k_{sv}R_{10})],$$

$$a_{76}^{(n)} = 2(\mu_s - i\omega\eta_s)n(n+1)[k_{sv}R_{10}h_n^{(2)\prime}(k_{sv}R_{10}) - h_n^{(2)}(k_{sv}R_{10})],$$

$$a_{77}^{(n)} = -2\eta_g(k_gR_{10})^2 j_n''(k_gR_{10}) + \left(\frac{ic_g^2\rho_g}{\omega} + \xi_g - \frac{2}{3}\eta_g\right) \times (k_gR_{10})^2 j_n(k_gR_{10}),$$

$$a_{78}^{(n)} = 2\eta_g n(n+1)[j_n(k_{gv}R_{10}) - k_{gv}R_{10}j_n'(k_{gv}R_{10})],$$

$$a_{81}^{(n)} = 0, \quad a_{82}^{(n)} = 0,$$

$$a_{83}^{(n)} = 2(i\omega\eta_s - \mu_s)[k_sR_{10}h_n^{(1)\prime}(k_sR_{10}) - h_n^{(1)}(k_sR_{10})],$$

$$a_{84}^{(n)} = 2(i\omega\eta_s - \mu_s)[k_sR_{10}h_n^{(2)\prime}(k_sR_{10}) - h_n^{(2)}(k_sR_{10})],$$

$$a_{85}^{(n)} = (i\omega\eta_s - \mu_s)\{(k_{sv}R_{10})^2 h_n^{(1)\prime\prime}(k_{sv}R_{10}) + [n(n+1) - 2]h_n^{(1)}(k_{sv}R_{10})\},$$

$$a_{86}^{(n)} = (i\omega\eta_s - \mu_s)\{(k_{sv}R_{10})^2 h_n^{(2)\prime\prime}(k_{sv}R_{10}) + [n(n+1) - 2]h_n^{(2)}(k_{sv}R_{10})\},$$

$$a_{87}^{(n)} = 2\eta_g[k_gR_{10}j_n'(k_gR_{10}) - j_n(k_gR_{10})],$$

$$a_{88}^{(n)} = \eta_g\{(k_{gv}R_{10})^2 j_n''(k_{gv}R_{10}) + [n(n+1) - 2]j_n(k_{gv}R_{10})\}.$$

The elements of the matrix $\boldsymbol{B}_0$ $(n=0)$ are given by

$$b_1^{(0)} = k_lR_{20}j_1(k_lR_{20}), \quad b_2^{(0)} = 0,$$

$$b_3^{(0)} = 2\eta_l(k_lR_{20})^2 j_1'(k_lR_{20}) + \frac{2i\sigma_l}{\omega R_{20}}k_lR_{20}j_1(k_lR_{20}) + \left(\frac{ic_l^2\rho_l}{\omega} + \xi_l - \frac{2}{3}\eta_l\right)(k_lR_{20})^2 j_0(k_lR_{20}),$$

$$b_4^{(0)} = 0.$$

The elements of the matrix $\boldsymbol{B}$ $(n \geq 1)$ are given by

$$b_1^{(n)} = -k_lR_{20}j_n'(k_lR_{20}), \quad b_2^{(n)} = -j_n(k_lR_{20}),$$

$$b_3^{(n)} = 0, \quad b_4^{(n)} = 0,$$

$$b_5^{(n)} = -2\eta_l(k_lR_{20})^2 j_n''(k_lR_{20}) + \frac{i\sigma_l(n-1)(n+2)}{\omega R_{20}}k_lR_{20}j_n'(k_lR_{20}) + \left(\frac{ic_l^2\rho_l}{\omega} + \xi_l - \frac{2}{3}\eta_l\right)(k_lR_{20})^2 j_n(k_lR_{20}),$$

$$b_6^{(n)} = 2\eta_l[j_n(k_lR_{20}) - k_lR_{20}j_n'(k_lR_{20})],$$

$$b_7^{(n)} = 0, \quad b_8^{(n)} = 0.$$

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

Davidson, B. J., and Riley, N. (**1971**). "Cavitation microstreaming," J. Sound Vib. **15**, 217–233.

Doinikov, A. A. (**1994**). "Acoustic radiation pressure on a compressible sphere in a viscous fluid," J. Fluid Mech. **267**, 1–21.

Doinikov, A. A. (**1997**). "Acoustic radiation force on a spherical particle in a viscous heat-conducting fluid. I. General formula," J. Acoust. Soc. Am. **101**, 713–721.

Doinikov, A. A., and Bouakaz, A. (**2010**). "Acoustic microstreaming around a gas bubble," J. Acoust. Soc. Am. **127**, 703–709 (**2010**).

Elder, S. A. (**1959**). "Cavitation microstreaming," J. Acoust. Soc. Am. **31**, 54–64.

Jackson, H. W., Barmatz, M., and Shipley, C. (**1988**). "Equilibrium shape and location of a liquid drop acoustically positioned in a resonant rectangular chamber," J. Acoust. Soc. Am. **84**, 1845–1862.

Landau, L. D., and Lifshitz, E. M. (**1986**). *Theory of Elasticity* (Pergamon, Oxford).

Landau, L. D., and Lifshitz, E. M. (**1987**). *Fluid Mechanics* (Pergamon, Oxford).

Lee, C. P., and Wang, T. G. (**1990**). "Outer acoustic streaming," J. Acoust. Soc. Am. **88**, 2367–2375.

Liu, X., and Wu, J. (**2009**). "Acoustic microstreaming around an isolated encapsulated microbubble," J. Acoust. Soc. Am. **125**, 1319–1330.

Longuet-Higgins, M. S. (**1998**). "Viscous streaming from an oscillating spherical bubble," Proc. R. Soc. London, Ser. A **454**, 725–742.

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.

Wu, J., and Du, G. (**1997**). "Streaming generated by a bubble in an ultrasound field," J. Acoust. Soc. Am. **101**, 1899–1907.

Modulation of ultrasound to produce multifrequency radiation force[a)]

Matthew W. Urban,[b)] Mostafa Fatemi, and James F. Greenleaf
Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, Minnesota 55905

(Received 24 August 2009; revised 15 December 2009; accepted 26 December 2009)

Dynamic radiation force has been used in several types of applications, and is performed by modulating ultrasound with different methods. By modulating ultrasound, energy can be transmitted to tissue, in this case a dynamic force to elicit a low frequency cyclic displacement to inspect the material properties of the tissue. In this paper, different types of modulation are explored including amplitude modulation (AM), double sideband suppressed carrier amplitude modulation AM, linear frequency modulation, and frequency-shift keying. Generalized theory is presented for computing the radiation force through the short-term time average of the energy density for these various types of modulation. Examples of modulation with different types of signals including sine waves, square waves, and triangle waves are shown. Using different modulating signals, multifrequency radiation force with different numbers of frequency components can be created, and can be used to characterize tissue mimicking materials and soft tissue. Results for characterization of gelatin phantoms using a method of vibrating an embedded sphere are presented. Different degrees of accuracy were achieved using different modulation techniques and modulating signals. Modulating ultrasound is a very flexible technique to produce radiation force with multiple frequency components that can be used for various applications.
 [DOI: 10.1121/1.3294487]

PACS number(s): 43.25.Qp, 43.35.Yb, 43.35.Mr [TDM] Pages: 1228–1238

I. INTRODUCTION

Ultrasound radiation force is the phenomenon of acoustic waves transferring their momentum upon striking or interacting with an object thus creating a force and imparting motion to the object.[1,2] Investigation of ultrasound radiation force and its applications has been an area of research for almost 8 decades. Theoretical studies on radiation pressure were undertaken in Refs. 1–7. Within the last 15 years applications of radiation force in the medical imaging field have grown considerably making it an active area of research.

Ultrasound radiation force has been used to calibrate the output of transducers using the radiation force balance method.[8–10] The manipulation of spheres, drops, bubbles, and other objects using radiation force has also been explored.[11–16] Methods based on radiation force have been developed to characterize different materials such as soft gelatin phantoms and soft tissue.[17–19]

One of the emerging uses of ultrasound radiation force is its use in ultrasound-based elasticity imaging. Focused ultrasound is used to exert a force on soft tissue and the mechanical response is measured with ultrasound-based or other methods.[19–25]

Radiation force can be classified as either static or dynamic.[26,27] Static radiation force is produced using continuous wave ultrasound to exert a constant force. Dynamic radiation force is created using ultrasound that has some sort of amplitude modulation through time.[22,27] The force changes with time based on the modulating function.

Dynamic ultrasound radiation force can be subdivided further into three categories: quasi-static, amplitude modulated, and frequency modulated. The quasi-static case results from using a finite duration toneburst of ultrasound, which exerts a transient force while the ultrasound is transmitted. This type of radiation force is used in acoustic radiation force impulse (ARFI) imaging where short tonebursts of ultrasound are used to impulsively excite tissue and the response is measured with ultrasound.[19]

Amplitude modulated ultrasound can be used to create a harmonic or multifrequency radiation force. This type of radiation force has been used in vibro-acoustography and shear wave elasticity imaging to displace tissue with a known harmonic function.[20,22,28,29] The use of amplitude modulation will be the main subject of this study and general cases will be described in more detail.

Lastly, frequency modulation (FM) of the ultrasound signal can be used to create a radiation force that changes frequency with time. Typically, the frequency shifts linearly through time and a "chirp" results. Chirped ultrasound has been used in radiation force applications in vibro-acoustography and in applications using contrast microbubbles to reduce standing wave artifacts.[30–32]

Modulation is used in communications to transmit information in one frequency band using a carrier signal that can be transmitted over large distances. In our case, we use ul-

[a)]Portions of this work were presented in "Modulated ultrasound and multifrequency radiation force," 154th Meeting of the Acoustical Society of America.

[b)]Author to whom correspondence should be addressed. Electronic mail: urban.matthew@mayo.edu

trasound at megahertz frequencies as a carrier signal to transfer energy, or a vibration force, at low frequencies, by modulating the ultrasound to elicit the desired stimulation force. We gain the advantage of beamforming at ultrasound frequencies for localizing the radiation force energy to create vibration motion at low frequencies such that the displacements are large enough to be measured with ultrasound or other techniques.

In this paper, we will present general theory that describes the formation of ultrasound radiation force using full amplitude modulation (AM) as well as double sideband suppressed carrier amplitude modulation (DSB-SC AM). The generation of radiation force with different modulating functions will be explored. Validation of the time- and frequency-domain relationships will be demonstrated with experimental data. Also, an example of material characterization using different modulation techniques and functions will be shown.

II. METHODS

The acoustic radiation force on an object can be defined as follows:

$$F(t) = d_r S \langle E(t) \rangle_T, \tag{1}$$

where $F(t)$ is the radiation force, d_r is the drag coefficient, S is the surface over which the ultrasound energy acts, and $\langle E(t) \rangle_T$ is the short-term time average of the energy density. The short-term time average is computed such that T, the averaging period, is longer than the period of the ultrasound wave but much shorter than the modulation period, $2\pi/\omega_c \ll T \ll 2\pi/\omega_m$ where ω_c and ω_m are the carrier (ultrasound) frequency and the modulation frequency, respectively.[22] If we have an absorbing medium we can write $F(t) = 2\alpha \langle I(t) \rangle_T / c = \alpha \langle E(t) \rangle_T$, where α is the ultrasound attenuation coefficient and $I(t)$ is the ultrasound intensity.[19] Assuming plane wave propagation, the energy density is defined as

$$E(t) = \frac{p^2(t)}{\rho c^2}. \tag{2}$$

Three types of modulation will be explored in this paper: amplitude modulation, double sideband suppressed carrier amplitude modulation, and frequency modulation. In this paper, we will ignore the spatial dependence of the focused ultrasound fields as that adds another dimension of complexity that is beyond the scope of this work.

A. AM

We start with amplitude modulation by defining $A(t)$ as the modulation function as follows, including a dc offset:[33]

$$A(t) = A_c[1 + \mu x_m(t)], \tag{3}$$

where A_c is the amplitude, μ is the modulation index that can vary from $0 \le \mu \le 1$, and $x_m(t)$, where $|x_m(t)| \le 1$, is the baseband or modulation function. The pressure for an AM ultrasound beam is given as

$$p(t) = A(t)\cos(\omega_c t) = A_c[1 + \mu x_m(t)]\cos(\omega_c t), \tag{4}$$

where ω_c is the carrier frequency, or in this case the ultrasound frequency. Equation (4) can be expanded as

$$p(t) = A_c \cos(\omega_c t) + A_c \mu x_m(t) \cos(\omega_c t). \tag{5}$$

To find the radiation force we need to square the pressure and take the short-term time average. We start by squaring the relationship in Eq. (5) and expand using the trigonometric identity $\cos^2(\omega t) = (1 + \cos(2\omega t))/2$.

$$p^2(t) = A_c^2 \cos^2(\omega_c t) + A_c^2 \mu^2 x_m^2(t) \cos^2(\omega_c t) + 2A_c^2 \mu x_m(t) \cos^2(\omega_c t), \tag{6}$$

$$p^2(t) = A_c^2 \left(\frac{1 + \cos(2\omega_c t)}{2} \right) + A_c^2 \mu^2 x_m^2(t) \left(\frac{1 + \cos(2\omega_c t)}{2} \right) + 2A_c^2 \mu x_m(t) \left(\frac{1 + \cos(2\omega_c t)}{2} \right). \tag{7}$$

After collecting terms, we arrive at the following expression for the square of the pressure function:

$$p^2(t) = \frac{A_c^2}{2}[\mu^2 x_m^2(t) + 2\mu x_m(t) + 1] + \frac{A_c^2}{2} \cos(2\omega_c t)[\mu^2 x_m^2(t) + 2\mu x_m(t) + 1]. \tag{8}$$

Applying the short-term time average eliminates the terms at $2\omega_c$ and the energy density is then computed in Eq. (10).

$$\langle p^2(t) \rangle_T = \frac{A_c^2}{2}[\mu^2 x_m^2(t) + 2\mu x_m(t) + 1], \tag{9}$$

$$\langle E(t) \rangle_T = \frac{A_c^2}{2\rho c^2}[\mu^2 x_m^2(t) + 2\mu x_m(t) + 1]. \tag{10}$$

For the case in which $x_m(t) = \cos(\omega_m t)$, we can substitute and rewrite Eq. (10) as

$$\langle E(t) \rangle_T = \frac{A_c^2}{2\rho c^2}[\mu^2 \cos^2(\omega_m t) + 2\mu \cos(\omega_m t) + 1]. \tag{11}$$

Applying the same trigonometric identity as above and simplifying give

$$\langle E(t) \rangle_T = \frac{A_c^2}{2\rho c^2}\left[\frac{\mu^2}{2}(1 + \cos(2\omega_m t)) + 2\mu \cos(\omega_m t) + 1 \right], \tag{12}$$

$$\langle E(t) \rangle_T = \frac{A_c^2}{4\rho c^2}[\mu^2 \cos(2\omega_m t) + 4\mu \cos(\omega_m t) + 2 + \mu^2]. \tag{13}$$

The result in Eq. (13) is interesting because the energy density has a dc term, a term at ω_m and at $2\omega_m$. In most cases $\mu = 1$, which represents 100% modulation, but if μ approaches 0, then the first term at $2\omega_m$ decreases.

B. DSB-SC AM

Now we will consider the case where the modulating function is defined without a dc offset[33]

$$A(t) = A_c \mu x_m(t). \tag{14}$$

We again consider the pressure signal

$$p(t) = A(t)\cos(\omega_c t) = A_c \mu x_m(t)\cos(\omega_c t). \tag{15}$$

Squaring the pressure gives

$$p^2(t) = A_c^2 \mu^2 x_m^2(t)\cos^2(\omega_c t), \tag{16}$$

and expanding and collecting terms give

$$p^2(t) = \frac{A_c^2 \mu^2}{2} x_m^2(t) + \frac{A_c^2 \mu^2}{2} x_m^2(t)\cos(2\omega_c t). \tag{17}$$

The short-term time average once again eliminates the term at $2\omega_c$ and the energy density is given in Eq. (19).

$$\langle p^2(t)\rangle_T = \frac{A_c^2 \mu^2}{2} x_m^2(t), \tag{18}$$

$$\langle E(t)\rangle_T = \frac{A_c^2 \mu^2}{2\rho c^2} x_m^2(t). \tag{19}$$

For the case in which $x_m(t)=\cos(\omega_m t)$, we can substitute and rewrite Eq. (19) as

$$\langle E(t)\rangle_T = \frac{A_c^2 \mu^2}{2\rho c^2}\cos^2(\omega_m t), \tag{20}$$

and then expanding we obtain

$$\langle E(t)\rangle_T = \frac{A_c^2 \mu^2}{4\rho c^2}(1 + \cos(2\omega_m t)). \tag{21}$$

This type of modulation provides a dc term and a component at $2\omega_m$, and in most cases $\mu=1$, which causes $\mu^2=1$.

As another more complex case, we can define $\mu=1$ and $x_m(t)$ as a square wave signal as has been previously explored,[34] but will be summarized here for emphasis.

$$x_m(t) = a\Pi\left(\frac{t - T_b/2}{T_b}\right) \otimes \frac{1}{T_r}\mathrm{III}\left(\frac{t}{T_r}\right). \tag{22}$$

Using Bracewell's notation and conventions, the radiation force can be written as a temporal convolution of an impulse train $\mathrm{III}(t)$, with a time-shifted rectangular function $\Pi(t)$.[35] The time-shifted rectangular function has width T_b, amplitude a, and the time shift is $T_b/2$. A repetition period of T_r, or a rate $f_r=1/T_r$, is used for the impulse train and $1/T_r$ is used to maintain unit height. The duty cycle can be written as $D = T_b/T_r = T_b f_r$.

Squaring $x_m(t)$ yields the same function assuming unit amplitude. We can examine the frequency components by evaluating the Fourier transform, where $X_m(f)$ is the Fourier transform of $x_m(t)$.

$$X_m(f) = \frac{aT_b}{T_r}\sum_{n=-\infty}^{\infty} e^{-i\pi T_b n/T_r}\,\mathrm{sinc}\left(n\frac{T_b}{T_r}\right)\delta\left(f - \frac{n}{T_r}\right). \tag{23}$$

We can substitute the repetition frequency f_r for the $1/T_r$ terms as follows:

$$X_m(f) = af_r T_b \sum_{n=-\infty}^{\infty} e^{-i\pi T_b n f_r}\,\mathrm{sinc}(T_b n f_r)\,\delta(f - nf_r). \tag{24}$$

We can further simplify by writing in terms of the duty factor.

$$X_m(f) = aD \sum_{n=-\infty}^{\infty} e^{-i\pi nD}\,\mathrm{sinc}(nD)\,\delta(f - nf_r). \tag{25}$$

The square wave radiation force has components at harmonics of the toneburst repetition frequency f_r, and the amplitude of those components will be governed by a sinc function that has the duty factor in its argument. The frequency-domain representation will be a set of impulses located at integral multiples of f_r and modulated by $af_r T_b e^{-i\pi T_b n f_r}\,\mathrm{sinc}(T_b n f_r)$.

C. FM

If we consider frequency modulation we can write the modulating function as[33]

$$A(t) = A_c(1 + \cos(\phi(t))). \tag{26}$$

In this paper, we will consider $\phi(t)$ to be a quadratic function with respect to time defined as $\phi(t)=((\omega_2-\omega_1)/(t_2-t_1))t^2 = (\Delta\omega/\Delta t)t^2$. The instantaneous frequency, which is the derivative of $\phi(t)$, is then a linear function $d\phi(t)/dt = 2(\Delta\omega/\Delta t)t$. The pressure is defined as

$$p(t) = A(t)\cos(\omega_c t) = A_c\left(1 + \cos\left(\frac{\Delta\omega}{\Delta t}t^2\right)\right)\cos(\omega_c t). \tag{27}$$

We then square the pressure and take the short-term time average and then arrive at the energy density.

$$\begin{aligned} p^2(t) &= A_c^2\left(1 + \cos\left(\frac{\Delta\omega}{\Delta t}t^2\right)\right)^2\cos^2(\omega_c t) \\ &= A_c^2\left[1 + 2\cos\left(\frac{\Delta\omega}{\Delta t}t^2\right) + \left(1 + \cos\left(2\frac{\Delta\omega}{\Delta t}t^2\right)\right)/2\right] \\ &\quad \times\left[\frac{(1 + \cos(2\omega_c t))}{2}\right], \end{aligned} \tag{28}$$

$$\langle p^2(t)\rangle_T = \frac{3A_c^2}{4} + A_c^2\cos\left(\frac{\Delta\omega}{\Delta t}t^2\right) + \frac{A_c^2}{4}\cos\left(2\frac{\Delta\omega}{\Delta t}t^2\right), \tag{29}$$

$$\langle E(t)\rangle_T = \frac{3A_c^2}{4\rho c^2} + \frac{A_c^2}{\rho c^2}\cos\left(\frac{\Delta\omega}{\Delta t}t^2\right) + \frac{A_c^2}{4\rho c^2}\cos\left(2\frac{\Delta\omega}{\Delta t}t^2\right). \tag{30}$$

If DSB-SC AM is used, $A(t)=A_c\cos(\phi(t))$, and $\langle E(t)\rangle_t$ is defined as

$$\langle E(t)\rangle_T = \frac{3A_c^2}{4\rho c^2} + \frac{A_c^2}{4\rho c^2}\cos\left(2\frac{\Delta\omega}{\Delta t}t^2\right). \tag{31}$$

D. Frequency-shift keying (FSK)

As an example of encoding the radiation force we present an example using FSK. The principle of this modulation technique is that the frequency of the carrier or the

modulating signal is adjusted to binary bits, 0 and 1.[33] In this study, we change the frequency of the modulating signal. To accomplish this, traditional AM or DSB-SC AM could be used and the modulating signal would change based whether the bit in a code $s[n]$ was 0 or 1. The energy density could be defined as in the following equation using AM where $x_{m,0}(t)$ is used to encode $s[n]$ as 0 and $x_{m,1}(t)$ is used to encode $s[n]$ as 1.

$$\langle E(t)\rangle_T = \begin{cases} \frac{A_c^2}{2\rho c^2}[x_{m,0}^2(t) + 2x_{m,0}(t) + 1], & s[n] = 0 \\ \frac{A_c^2}{2\rho c^2}[x_{m,1}^2(t) + 2x_{m,1}(t) + 1], & s[n] = 1. \end{cases} \quad (32)$$

III. EXPERIMENTS

We will present experiments designed to validate the theory that has been presented above and give an example of an application using multifrequency radiation force similar to that presented in Ref. 28. We modulated an ultrasound signal with different functions and derived the radiation force with the pressure from the needle hydrophone measurements. Because the force is a quadratic function of the pressure, we could not directly validate the force. We approached this validation in an indirect fashion.

To validate the theory presented in this study and exhibit an application of this general technique we measured the motion of a steel sphere in a gelatin phantom as we changed the modulating signal used. This is a variation of the method introduced by Chen *et al.*[17] to measure the viscoelastic material properties of the gelatin by finding the frequency response of the sphere induced by radiation force. First, we will use the method described by Chen *et al.*[17] to characterize the viscoelastic material properties of the gelatin. We will then use the theory presented below to compute the sphere displacement from the force derived from the needle hydrophone measurements, and the theory that was presented above. We will compare that displacement signal with the displacement we measure.

There is also another more indirect way of validating the theory that was presented above. If we use the same phantom and change the modulating signal, we will induce different motion. Using the derived force from needle hydrophone measurements and the motion measurements, we can find the material properties of the gelatin. If the developed theory is correct, then we should obtain the same values for the material properties regardless of the modulation technique.

These measurements with modulated ultrasound signals will provide an application of this general method for the characterization of the viscoelastic material properties in one measurement as opposed to the original method put forth by Chen *et al.*,[17] where multiple single frequency measurements are performed.

A. Needle hydrophone measurements

Needle hydrophone measurements were performed to measure the ultrasound radiofrequency (rf) signal in a water tank. This signal was used to derive the radiation force time- and frequency-domain representations to compare with results from the analytic expressions detailed above. These measurements will be used as a forcing function input for later measurements made in a gelatin sphere phantom to evaluate the transfer function of the phantom material. A 0.5 mm diameter needle hydrophone (NTR Systems, Seattle, WA) was positioned in a large water tank at the focus of a 3.0 MHz transducer with a diameter of 45 mm and a spherical focus of 70 mm, which was assembled in-house using a focused piezoelectric crystal (Boston Piezo-Optics, Inc., Bellingham, MA).

Different waveforms and types of modulation were used for validation of the theory presented above. A sine wave of frequency 100 Hz was used for $x_m(t)$ for AM, and a sine wave of frequency 50 Hz was used for $x_m(t)$ for DSB-SC AM. A square wave with a 50% duty cycle, T_b=5 ms and T_r=10 ms, was used with amplitude modulation. A triangle wave was used with AM. An AM implementation with a linear FM signal was used with a bandwidth of 1 μHz–5000 Hz, and the frequency sweep was performed in 50 ms. Lastly, a FSK signal was formed by defining a digital zero as 8 cycles of a square wave with 50% duty cycle with T_r of 5 ms (f_r=200 Hz) and a digital one as 4 cycles of a square wave with 50% duty cycle with T_r=10 ms (f_r =100 Hz).

B. Steel sphere phantom study

From the ultrasound rf signals, the radiation force can be calculated, within a constant multiplier, by squaring the pressure and applying a low-pass filter for each modulation waveform. We used the same waveforms as were characterized in the needle hydrophone measurements for vibrating a 440C stainless steel sphere with diameter of 1.59 mm embedded in three types of gelatin phantoms. Phantoms 1, 2, and 3 were made from 300 Bloom gelatin powder (Sigma-Aldrich, St. Louis, MO) with concentrations of 10%, 10%, and 12% by volume, respectively, and glycerol (Sigma-Aldrich, St. Louis, MO) with concentrations of 0%, 10%, and 10% by volume, respectively. A preservative of potassium sorbate (Sigma-Aldrich, St. Louis, MO) was also added with a concentration of 10 g/L to all three phantoms. Unless otherwise noted, the results shown are from phantom 1. A Doppler laser vibrometer (Polytec, Waldbronn, Germany) was used to measure the velocity of the sphere. We measure the velocity because it is proportional to the force imparted on the sphere.

The velocity can be used to characterize the system if the force is known. We can apply linear systems modeling where $v(t)$ is the velocity, $f(t)$ is the force, and $z(t)$ is the mechanical impedance function for the sphere/gelatin phantom. The force is equal to the convolution of the velocity and the impedance function.

$$f(t) = v(t) \otimes z(t). \quad (33)$$

In the frequency domain the force denoted as $F(\omega)$ is the product of the frequency-domain representations of the velocity $V(\omega)$ and the impedance $Z(\omega)$.[17,36]

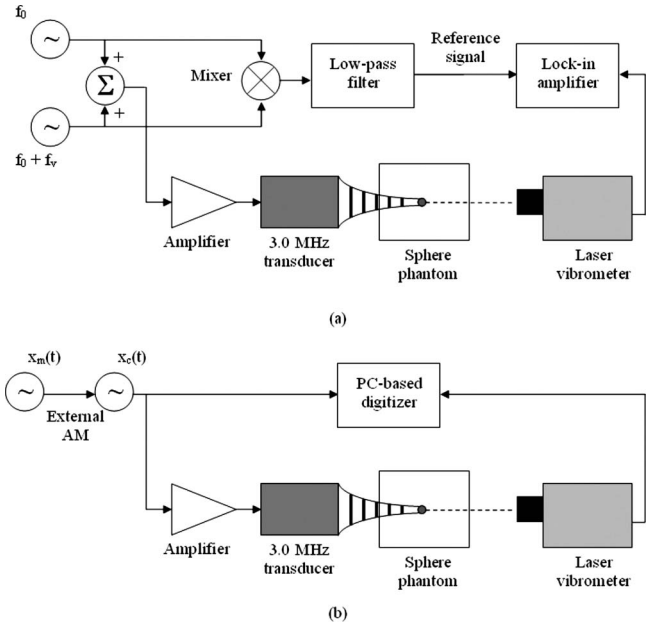


FIG. 1. Experimental setups. (a) Experimental setup with DSB-SC AM modulation, sphere velocity was measured with the laser vibrometer and lock-in filter. (b) Experimental setup with modulated ultrasound, sphere velocity was measured with the laser vibrometer and recorded with PC-based digitizer.

$$F(\omega) = V(\omega)Z(\omega). \tag{34}$$

We can estimate the impedance function $\tilde{Z}(\omega)$ by dividing the force function derived from the transmitted ultrasound measured before amplification by the measured velocity

$$\tilde{Z}(\omega) = \frac{F(\omega)}{V(\omega)}. \tag{35}$$

Using a model defined by Ostreicher and utilized by Chen *et al.*, we can find the shear elasticity μ_1 and shear viscosity μ_2, assuming a value of 1000 kg/m^3 for the density of the gelatin.[17,37] We consider this method to be our gold standard, or reference, measurement for μ_1 and μ_2.

To obtain a reference measurement for characterization of the gelatin, a DSB-SC AM signal with $x_m(t)=\cos(2\pi f_m t)$ where f_m was varied from 25 to 1000 Hz to yield vibration frequencies of 50–2000 Hz. The output of the laser vibrometer was processed by a lock-in amplifier (Signal Recovery, Oak Ridge, TN) to obtain the measurements of the magnitude of the velocity. The advantage of using the modulated ultrasound waveforms is that they produce multiple frequencies so that the impedance function can be estimated in one measurement compared to making a set of discrete measurements at single frequencies with the lock-in filter. However, when using modulation, each frequency component has a different magnitude and so we must take that into account when trying to characterize the mechanical system as indicated in Eq. (35). We normalize for the magnitude variation of the force when we use the DSB-SC AM method, varying the frequency to one value at a time, because we ensure that the same force magnitude is used for each measurement. A figure of the experimental setup with the lock-in filter is shown in Fig. 1(a) and with the digitizer in Fig. 1(b).

For the experiment with the lock-in filter we use two signal generators (33250A, Agilent, Santa Clara, CA) to generate signals at f_0 and f_0+f_v where f_0 is 3.0 MHz and f_v is the desired vibration frequency, and the carrier frequency is $f_c=(f_0+f_v)/2$. Those signals are added using a hybrid junction (M/A COM, Inc., Lowell, MA) and used as input to the amplifier. The amplified output was applied to the 3.0 MHz transducer. The ultrasound signals were also used as inputs for a mixer. The output of the mixer was low-pass filtered to get a reference signal at f_v for the lock-in amplifier (Signal Recovery, Oak Ridge, TN). The laser vibrometer output was also input to the lock-in amplifier.

For the experiment with the PC-based digitizer (Alazartech, Montreal, QC, Canada), the modulating signal $x_m(t)$ was created using one signal generator (33250A, Agilent, Santa Clara, CA) and that signal was applied to the external AM input of another signal generator set at 100% modulation index that was generating the 3.0 MHz ultrasound carrier frequency $x_c(t)$. The modulated signal was sent to one channel of the digitizer and also amplified and applied to the transducer. The laser vibrometer signal was sent directly to the digitizer. The sampling frequency for both signals was 50 MHz.

We fit the velocity data from the laser vibrometer to the model first presented by Ostreicher and subsequently used by Chen *et al.*, to characterize tissue mimicking gelatin[17,37] to find the viscoelastic parameters μ_1 and μ_2.

$$V(\omega) = \frac{F(\omega)}{Z(\omega)} = \frac{F(\omega)}{Z_r(\omega) + Z_m(\omega)}. \tag{36}$$

The impedance terms, derived under plane wave conditions, are given as

$$Z_r(\omega) = -i\frac{4\pi a^3}{3}\rho\omega \times \frac{\left(1-\frac{3i}{ah}-\frac{3}{a^2h^2}\right)-2\left(\frac{i}{ah}+\frac{1}{a^2h^2}\right)\left(3-\frac{a^2k^2}{aki+1}\right)}{\left(\frac{i}{ah}+\frac{1}{a^2h^2}\right)\frac{a^2k^2}{aki+1}+\left(2-\frac{a^2k^2}{aki+1}\right)}, \tag{37}$$

$$Z_m(\omega) = -i\frac{4\pi a^3}{3}\rho_s\omega, \tag{38}$$

where $k=\sqrt{\rho\omega^2/(2\mu+\lambda)}$, $h=\sqrt{\rho\omega^2/\mu}$, $\mu=\mu_1+i\omega\mu_2$, $\lambda=\lambda_1+i\omega\lambda_2$, a is the radius of the sphere, ρ is the mass density of the medium, μ_1 and μ_2 are the shear elasticity and viscosity of the medium, λ_1 and λ_2 are the bulk elasticity and viscosity, and ρ_s is the mass density of the sphere. The values used for this model implementation were $\rho=1000$ kg/m^3, $a=0.794$ mm, $\rho_s=7849$ kg/m^3, $\lambda_1=2.48$ GPa, and $\lambda_2=0$. The curve fits to the model were performed using the nlinfit.m function in MATLAB (The Mathworks, Natick, MA).

To validate the theory on modulating ultrasound with the measurements, we examine the displacement of the sphere from theoretical and experimental bases. For the sphere in the gelatin phantom we can derive the displacement by inte-

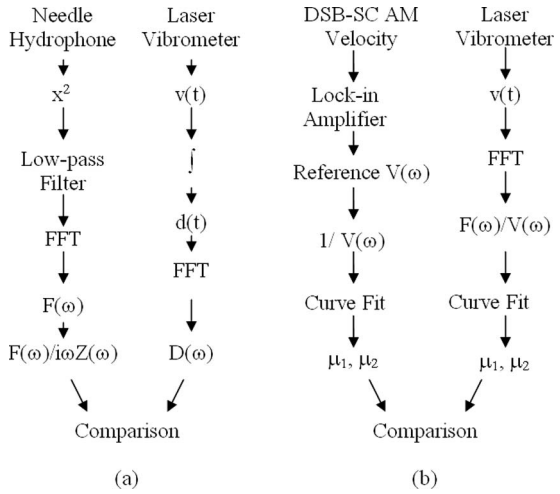


FIG. 2. Validation methods for the multifrequency radiation force theory. (a) The force is derived from the pressure measured by the needle hydrophone. The pressure is squared and low-pass filtered and the spectrum is computed. The force spectrum is divided by $i\omega Z(\omega)$. The velocity measured by the laser vibrometer was integrated to get displacement. The displacement spectrum was computed and compared to the signal derived from the needle hydrophone. (b) The DSB-SC AM measurements made with the lock-in amplifier provide a reference measurement of impedance $Z(\omega)$, and a curve fitting was performed to get μ_1 and μ_2. The laser vibrometer velocity signal is measured and the reference force signal is divided by $V(\omega)$. The curve fit is performed and μ_1 and μ_2 values are compared.

grating Eq. (36). In the frequency domain, this is accomplished by dividing both sides of the equation by $i\omega$, yielding

$$D(\omega) = \frac{V(\omega)}{i\omega} = \frac{F(\omega)}{i\omega(Z_r(\omega) + Z_m(\omega))}. \tag{39}$$

We can take the force that we derive from the pressure measurements using the needle hydrophone and divide by the product $i\omega Z(\omega)$ that is computed from the results of the reference measurement using single modulation frequencies and the lock-in amplifier. We then will compare this with the measured displacement. If the theory is correct, the results should match. A flow diagram of this process is shown in Fig. 2(a).

We also used a large set of modulating waveforms to estimate the impedance function for the gelatin phantom including AM with square waves of duty cycles ranging from 25% to 80%, a triangle wave, a sawtooth wave, a linear FM signal, and FSK performed with the square waves of duty cycles ranging from 25% to 75%, a triangle wave, and a sawtooth wave. Using DSB-SC AM we used a triangle wave, a linear FM signal, and a FSK with a triangle wave. These modulating signals yielded a multifrequency radiation force and we evaluated how these types of signals could be used to measure the mechanical properties of the gelatin phantom in one measurement. The reference measurements and the measurements with different modulating waveforms were performed on all three phantoms that were made for this study. A flow diagram for this process with the modulation functions is shown in Fig. 2(b).

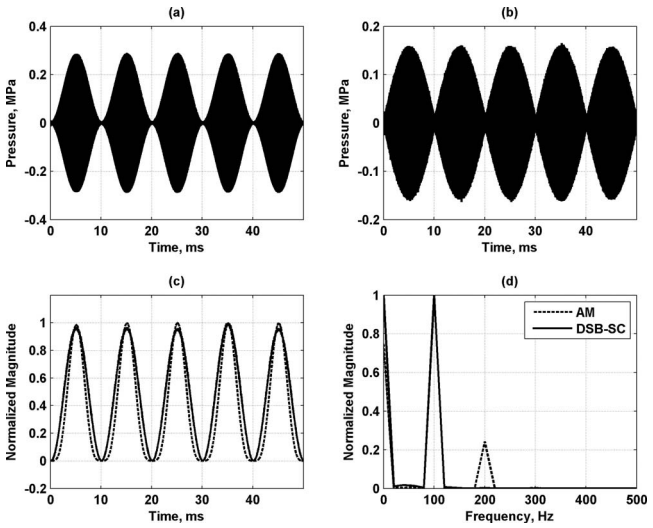


FIG. 3. Modulation with sinusoidal functions. (a) Ultrasound pressure for AM with sinusoidal function with frequency 100 Hz. (b) Ultrasound pressure for DSB-SC AM with sinusoidal function with frequency 50 Hz. (c) Calculated radiation force for modulation in (a) and (b). (d) Spectra for radiation force in (c). The legend in (d) applies to (c) and (d).

IV. RESULTS

A. Needle hydrophone measurements

Figure 3(a) shows the ultrasound pressure for AM performed with a sine wave with frequency of 100 Hz, and Fig. 3(b) shows the ultrasound pressure for DSB-SC AM performed with a sine wave with frequency of 50 Hz. The waveforms appear filled because of the many ultrasound cycles within the 50 ms timeframe of the transmission. Figure 3(c) shows the calculated radiation force produced by the two types of modulated ultrasound in Figs. 3(a) and 3(b). Figure 3(d) shows the frequency-domain representations of the signals in Fig. 3(c). The frequency components are in agreement in proportion and frequency as the analytic expressions in Eqs. (13) and (21) for the AM and DSB-SC AM cases, respectively.

Figure 4(a) shows the ultrasound pressure for AM with a square wave with 50% duty cycle. Figures 4(b) and 4(c) show the time- and frequency-domain representations of the intensity, which is proportional to the radiation force. The square wave has components at odd harmonics of the fundamental, 100 Hz. The components at the even harmonics are low compared to adjacent odd harmonics and are probably introduced because of the ultrasound overshoot at the beginning of a toneburst, and the extra transmission or standing wave at the end of the toneburst observed in the needle hydrophone measurements.

Figure 5(a) shows a triangle wave used for modulation. Figures 5(b) shows the intensity when the triangle wave is used in AM, respectively. The frequency-domain representations of the intensity are shown in Fig. 5(c). The triangle wave modulation signal provides components at harmonics of the fundamental frequency, 100 Hz. The use of different modulating functions changes the amplitude of the components that they produce in the radiation force.

Figure 6(a) shows a linear FM signal, and Fig. 6(b) shows the intensity. The frequency-domain representations

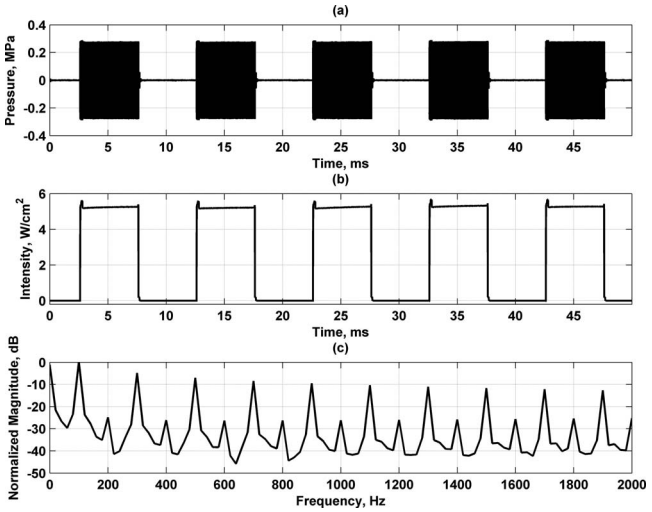


FIG. 4. Modulation with AM square wave with 50% duty cycle and frequency of 100 Hz. (a) Modulated ultrasound, (b) Intensity for modulated ultrasound in (a), and (c) spectrum for intensity in (b).

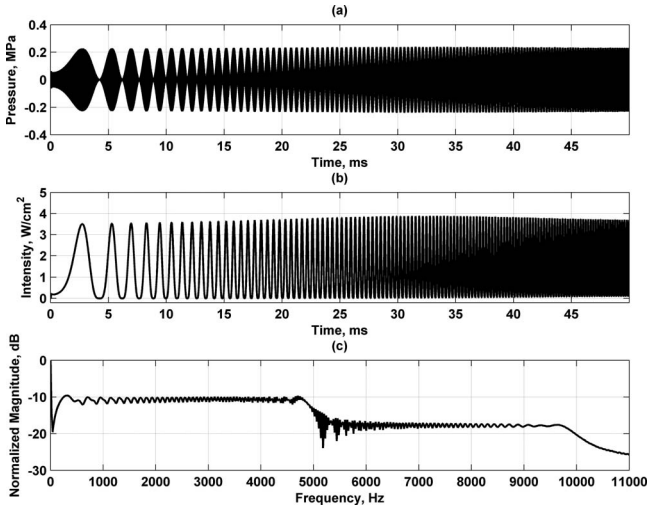


FIG. 6. Modulation with AM with linear FM signal that sweeps from 1 μHz to 5000 Hz in 50 ms. (a) Modulated ultrasound, (b) intensity for modulated ultrasound in (a), and (c) spectrum for intensity in (b).

of the FM function and the intensity are shown in Fig. 6(c). The use of frequency modulation provides a near constant bilevel radiation force from 0 to 5 kHz and from 5 to 10 kHz, which is separated by about 6 dB. However, the force is not constant at the extremes of the bandwidth used, particularly at the lower frequencies from 0 to 300 Hz and at the transition around 5 kHz.

For the case of the FSK with a square wave with two different fundamental frequencies, Fig. 7(a) shows the modulated ultrasound pressure signal. The intensity is shown in Fig. 7(b), and the frequency-domain representation of the intensity is shown in Fig. 7(c). The FSK square wave yields components at the odd harmonics of the fundamental frequencies of the modulating signals used to encode the different bits, 200 and 100 Hz, respectively. From the first segment (0–50 ms), components at 200, 600, 1000, 1400, and 1800 Hz result, and from the second segment (50–100 ms), components at 100, 300, 500, 700, 900, 1100, 1300, 1500, 1700, and 1900 result. In total, using this 2 bit code yields 15 unique components.

B. Sphere phantom characterization

We measured the velocity of a sphere induced by the multifrequency radiation force for different modulating functions directly with the laser vibrometer and derived the displacement by integrating the velocity signal. Figures 8(a) and 8(b) show the sphere displacement and velocity for the AM square wave excitation. Figures 8(c) and 8(d) show the sphere displacement and velocity for the AM triangle waveform. Figures 8(e) and 8(f) show the sphere displacement and velocity for the FSK square wave signal. The displacements look very similar to the force depicted in Figs. 4(b), 5(b), and 7(b) for the square, triangle, and FSK modulation functions. There is some very low frequency drift that is

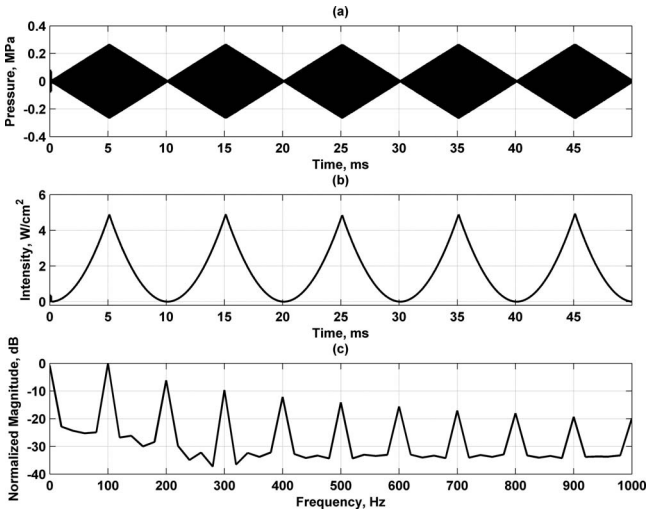


FIG. 5. Modulation with AM triangle wave with frequency of 100 Hz. (a) Modulated ultrasound, (b) intensity for modulated ultrasound in (a), and (c) spectrum for intensity in (b).

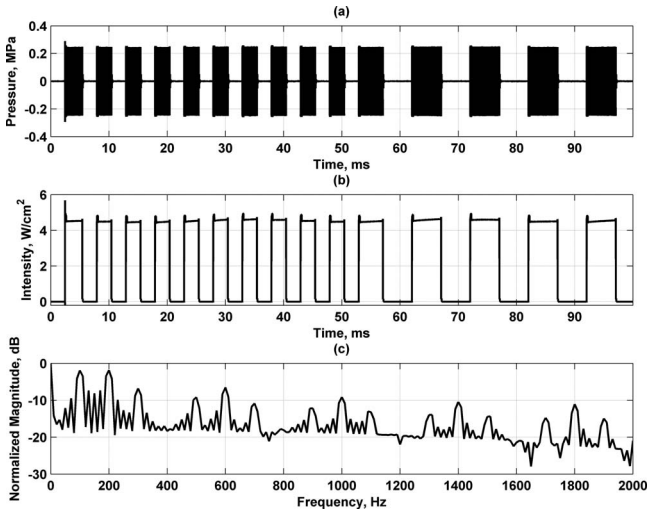


FIG. 7. Modulation with AM using FSK with square wave frequencies of 200 Hz for 0–50 ms and 100 Hz for 50–100 ms. (a) Modulated ultrasound, (b) intensity for modulated ultrasound in (a), and (c) spectrum for entire intensity in (b).

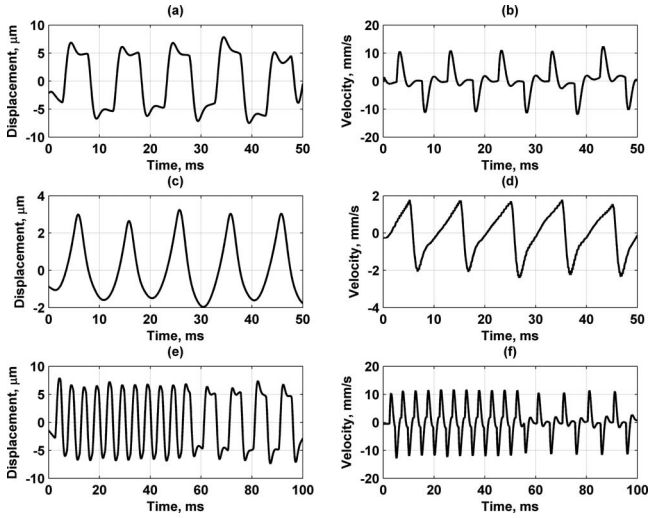


FIG. 8. Displacement and velocity of sphere embedded in gelatin phantom obtained using different modulation signals to produce radiation force. (a) Sphere displacement for AM square wave used in Fig. 4. (b) Sphere velocity for AM square wave. (c) Sphere displacement for AM triangle wave used in Fig. 5. (d) Sphere velocity for AM triangle wave. (e) Sphere displacement for FSK square wave used in Fig. 7. (f) Sphere velocity for FSK square wave.

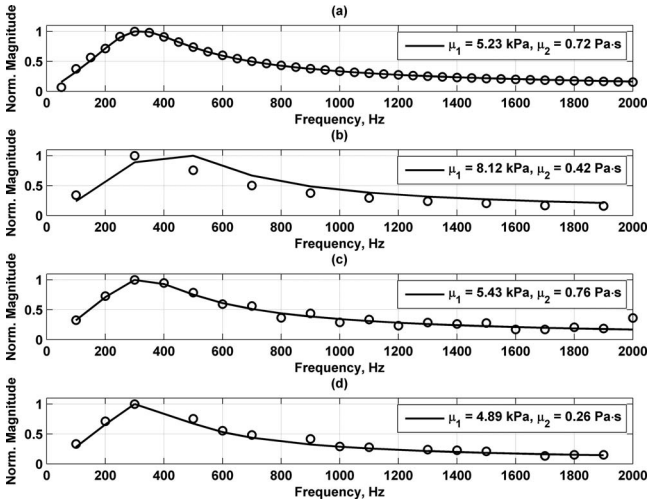


FIG. 10. Sphere phantom impedance functions and model fits for different modulation functions. (a) Data, denoted by the open circles, and model fit, denoted as the solid line, for DSB-SC AM with sinusoidal modulating function. The legends contain the material properties from the model fit in each respective panel, (b) Data and model fit for AM with square wave with 50% duty cycle as modulating function. (c) Data and model fit for AM with triangle wave as modulating function. (d) Data and model fit for AM with FSK square wave with 50% duty cycle as modulating function.

most prevalent in the AM triangle wave results. These qualitative similarities are an indication that the theory used to derive the force is correct.

The reference measurement with single frequencies was carried out and yielded $\mu_1=5.23$ kPa and $\mu_2=0.72$ Pa s for phantom 1. These values were used for computing the displacement using Eq. (39) and the velocity using Eq. (36) assuming a constant force for all frequencies. The normalized magnitude spectra for the displacement and velocity are shown in Fig. 9(a). If we look at these spectra from a linear systems perspective, the displacement has a low-pass filter characteristic and the velocity has a bandpass filter characteristic. The low-pass filter nature of the displacement response agrees with the qualitative assessment of the measured displacement previously mentioned and displayed in Fig. 8. We then took the derived force for square wave modulation from Fig. 4(b) and divided it by the computed product $i\omega Z(\omega)$ and compared that with the measurement from square wave modulation shown in Fig. 8(a). The normalized theoretical and measured results are shown in Fig. 9(b). The agreement is very good. This result serves as verification of the modulation theory.

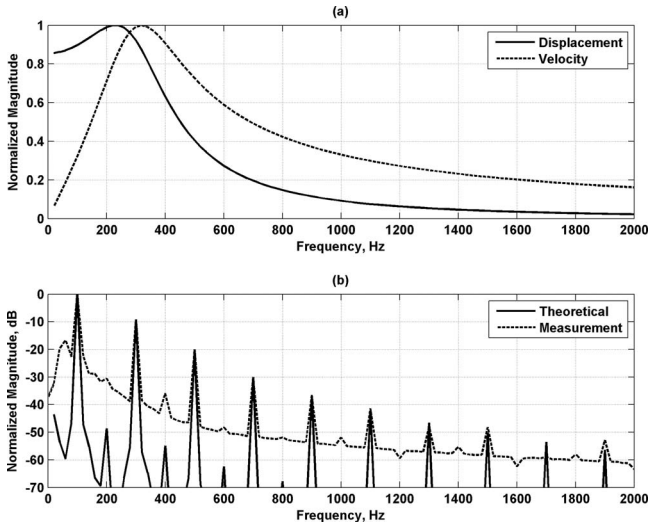


FIG. 9. Validation results using measured and calculated displacements. (a) Calculated displacement and velocity responses assuming a constant force at all frequencies. (b) Theoretical and measured displacement magnitude spectra results. The theoretical magnitude spectrum was calculated using the force derived from pressure measurements made with a needle hydrophone [Fig. 4(b)] divided by the calculated product $i\omega Z(\omega)$. The measured magnitude spectrum is from the displacement in Fig. 8(a).

Figure 10(a) shows the reference measurement made with the DSB-SC AM measurements and the estimated impedance function. Figure 10(b) shows the magnitude spectrum of the sphere phantom impedance function derived from the transmitted ultrasound and the velocity measured by the laser for the AM square waveform. Figures 10(c) and 10(d) show the magnitude spectra for the impedance function obtained using the AM triangle modulating signal and the FSK square wave signal, respectively. The number of frequency components with high signal-to-noise ratio (SNR) varies based on the modulating signal. The solid lines in the plots in Fig. 10 are the curve fits to the model for the embedded sphere in a viscoelastic medium. We performed this for the different waveforms used in this study and tabulated the results in Table I.

Noting that each modulating signal produces different components with significant SNR in the radiation force, we attempted to account for this fact by performing linear interpolation and extrapolation over the frequency bandwidth that was used for the reference measurement, 50–2000 Hz. We performed the linear interpolation and fit the results to the model as was explained above, and those results are also

TABLE I. Model fitting results for viscoelastic properties of gelatin with different modulation signals.

Modulation signal	μ_1 (kPa)	μ_1 (kPa) (interpolation)	μ_2 (Pa S)	μ_2 (Pa S) (interpolation)
DSB-SC sine (reference)	5.23	5.23	0.72	0.72
AM square 25%	5.15	5.32	0.62	0.88
AM square 30%	4.94	5.19	0.51	0.76
AM square 50%	4.57	5.60	0.19	0.77
AM square 75%	5.77	5.37	0.37	0.83
AM square 80%	5.89	6.00	0.53	0.57
AM triangle	5.27	5.55	0.67	0.59
AM sawtooth	5.22	5.30	0.58	0.63
AM chirp 5 kHz	7.24	6.85	−0.08	0.40
AM FSK square 25%	5.17	5.29	0.54	0.63
AM FSK square 50%	5.45	5.18	0.45	0.80
AM FSK square 75%	5.04	5.18	0.82	0.85
AM FSK triangle	5.47	5.56	0.86	0.90
AM FSK sawtooth	5.36	5.62	0.69	0.66
DSB-SC triangle	5.46	5.57	0.80	0.85
DSB-SC chirp 5 kHz	9.35	9.28	0.57	0.32
DSB-SC FSK triangle	5.33	5.44	0.79	0.87

summarized in Table I. Using the same modulation signals as shown Fig. 10, linear interpolation was performed and the model fits are shown in Fig. 11.

The reference measurement with single frequencies was carried out on phantoms 2 and 3 and yielded μ_1=7.21 kPa and μ_2=0.52 Pa s, and μ_1=8.92 kPa and μ_2=0.67 Pa s, respectively. The phantoms had increased shear elasticity, but did not have increased shear viscosity. We compared the material property estimates made from measurements using different modulation waveforms such as AM with square wave with 50% duty cycle, AM with a sawtooth wave, AM FSK

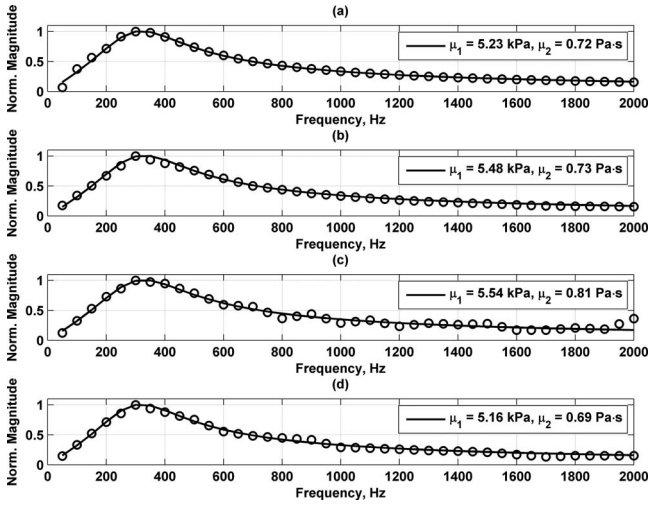


FIG. 11. Sphere phantom impedance functions and model fits for different modulation functions. (a) Data, denoted by the open circles, and model fit, denoted as the solid line, for DSB-SC AM with sinusoidal modulating function. The legends contain the material properties from the model fit in each respective panel. (b) Interpolated data and model fit for AM with square wave with 50% duty cycle as modulating function. (c) Interpolated data and model fit for AM with triangle wave as modulating function. (d) Interpolated data and model fit for AM with FSK square wave with 50% duty cycle as modulating function.

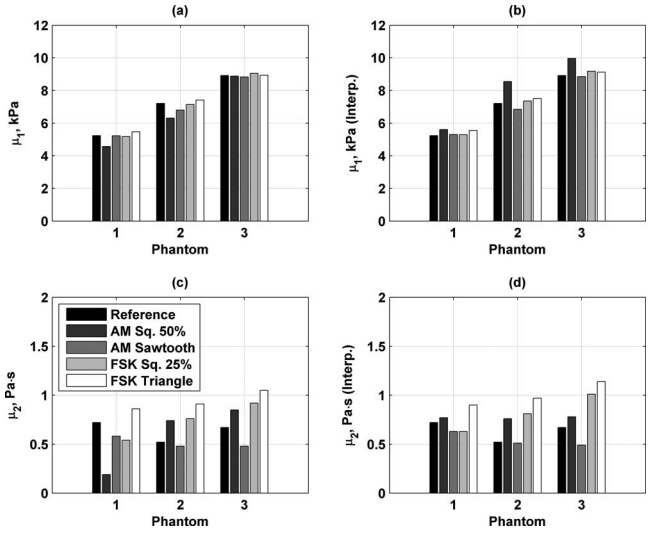


FIG. 12. Comparison of material property estimation for three gelatin phantoms for different modulation waveforms. The reference measurement is the DSB-SC AM measurements made with single frequencies. (a) Estimation of μ_1, (b) estimation of μ_1 with interpolation, (c) estimation of μ_2, and (d) estimation of μ_2 with interpolation.

with square wave with 25% duty cycle, and AM FSK with a triangle wave. Figure 12 summarizes those estimation results for all three phantoms for μ_1 and μ_2 without and with interpolation. In general, the estimates for μ_1 were similar to the reference measurement, but the estimates for μ_2 were more varied.

V. DISCUSSION

Using different types of modulation and modulating signals yields radiation force with different shapes and spectra. Using a sinusoidal modulation signal with either AM or DSB-SC AM yields spectra with different components and weighting of those components as predicted by the theory. When more complex modulating signals are used such as a square wave, a triangle wave, or a FM signal, the radiation force contains multiple frequency components. The magnitude of those components also changes and so different modulating signals could be used for different experiments based on the desired excitation.

We measured the cyclic velocity of a sphere induced by the multifrequency radiation force for different modulating functions. Using the velocity signals, we were able to estimate the viscoelastic properties of the gelatin in which the sphere was embedded. Figure 10 shows the fits using different modulation functions along with the reference measurement that was made using DSB-SC AM with a single sinusoid and varying the frequency. Using the triangle and FSK square waves seemed to produce results closest to the reference measurement, and the match was aided by the inclusion of more frequency components.

When linear interpolation was performed to achieve data with the same frequency resolution as the reference measurement, the fits matched better. The linear interpolation helped to compensate for frequency components that lacked significant SNR for certain modulating waveforms. For example, when using the square wave with 50% duty cycle [Fig.

10(b)], we only obtained components at 100, 300, 500,…, 1900 Hz, but the interpolation filled components at 200, 400, 600,…, 2000 Hz and improved the fitting. This was quantitatively evaluated and summarized in Table I. In a majority of cases, interpolation helped to improve the fits to obtain values of μ_1 and μ_2 that were closer to those of the reference measurement made at single frequencies. In general the values for μ_1 were in good agreement. The value of μ_2 was more difficult to estimate consistently, probably because the shear viscosity was fairly low. The worst results were given with the AM chirp signal and that may be due to low SNR at any given frequency during the course of the chirp as opposed to the other waveforms used. A slower chirp could be used and may improve the results because more time would be taken at a given frequency.

We compared the measured sphere displacement and that derived from the theoretical development, and the agreement was very good. We take this result as a validation of the theory. We also observed that different modulation signals yielded similar estimates of μ_1 and μ_2. Using the data in Table I, the median absolute percent errors in μ_1 were 4.12% and 3.94% without and with interpolation, respectively, and the median absolute percent errors in μ_2 were 21.54% and 16.56% without and with interpolation, respectively. The higher percent errors for μ_2 result because of the low value of μ_2. This agreement among material property estimates using different modulation signals would not be possible if the general theoretical basis was flawed. These results serve as an indirect validation as well as application of this multifrequency radiation force theory and method.

Modulating the ultrasound produces vibration with high SNR for a selected group of frequencies depending on the modulation function. The accuracy of the characterization of the viscoelastic properties of the gelatin material is dependent on the number of frequency components present and their respective weighting. Depending on the suspected material properties, one could choose one modulating signal versus another to optimize the ability to characterize the material. Either weighting at certain frequencies or desired bandwidth could be considered.

AM is probably the most versatile modulation technique while DSB-SC AM is also useful in creating modulated signals. FSK could be used for introducing codes into the radiation force if there was an application that necessitated this technique. FM could be used to distribute energy to a wide bandwidth. The modulation technique used for a given application depends on the way that the energy needs to be distributed in the forcing function.

One of the advantages of using modulation to produce a multifrequency radiation force is that one can obtain the same data in one measurement that the reference method took multiple measurements to obtain. The information gain can be quite high and the signal strength can be concentrated into known frequency components in the desired bandwidth through proper selection of the modulation function.

Two applications have arisen that utilize multifrequency radiation force to take advantage of obtaining information at multiple frequencies in one measurement. Vibro-acoustography has taken advantage of this multifrequency radiation force using DSB-SC AM waveforms to create multiple images from one scan.[28] Another method called shear wave dispersion ultrasound vibrometry (SDUV) uses square wave modulation to transmit multiple tonebursts of ultrasound at the frequency of the square wave f_r, and then elicits shear waves with components at harmonics of f_r.[25] The shear wave velocity is estimated at f_r and several of its harmonics to estimate the shear wave velocity dispersion and characterize the viscoelastic material properties of the tissue or medium.

One could also design modulating functions to emphasize certain frequency components with specific weightings. That signal could be entered as an arbitrary waveform in hardware to be used for the analog modulation function.

VI. CONCLUSION

In this paper we presented general theory for modulating ultrasound to produce multifrequency radiation force. We showed that using different modulation functions can create multifrequency radiation force with components at different frequencies and different relative weighting. As an example of the utility of using modulated ultrasound, we showed that this method could be used to characterize the viscoelastic properties of a tissue mimicking phantom material with differing degrees of accuracy based on the modulation function using *one* multifrequency measurement. Techniques such as SDUV and vibro-acoustography use these modulation techniques to perform noncontact, noninvasive characterization and imaging of materials and biological tissues.

ACKNOWLEDGMENTS

The authors are grateful to Randall Kinnick for hardware and experimental support, Tom Kinter for software and experimental support, and Jennifer Milliken for administrative support. M.W.U. is grateful to Dr. Shigao Chen for curve fitting procedure used for characterizing the gelatin phantoms and to Dr. Farid Mitri for helpful conversation. This work was supported in part by Grant Nos. EB002167, CA127325, CA91956, and CA121579 from the National Institutes of Health.

[1]P. J. Westervelt, "Acoustic radiation pressure," J. Acoust. Soc. Am. **29**, 26–29 (1957).
[2]G. R. Torr, "The acoustic radiation force," Am. J. Phys. **52**, 402–408 (1984).
[3]P. J. Westervelt, "The theory of steady forces caused by sound waves," J. Acoust. Soc. Am. **23**, 312–315 (1951).
[4]C. P. Lee and T. G. Wang, "Acoustic radiation pressure," J. Acoust. Soc. Am. **94**, 1099–1109 (1993).
[5]R. T. Beyer, "Radiation pressure—The history of a mislabeled tensor," J. Acoust. Soc. Am. **63**, 1025–1030 (1978).
[6]B.-T. Chu and R. E. Apfel, "Acoustic radiation pressure produced by a beam of sound," J. Acoust. Soc. Am. **72**, 1673–1687 (1982).
[7]Z. Y. Jiang and J. F. Greenleaf, "Acoustic radiation pressure in a three-dimensional lossy medium," J. Acoust. Soc. Am. **100**, 741–747 (1996).
[8]W. D. Shou, X. W. Huang, S. M. Duan, R. M. Xia, Z. L. Shi, X. M. Geng, and F. Q. Li, "Acoustic power measurement of high intensity focused ultrasound in medicine based on radiation force," Ultrasonics **44**, e17–e20 (2006).
[9]S. E. Fick and F. R. Breckenridge, "Ultrasonic power output measurement by pulsed radiation pressure," J. Res. Natl. Inst. Stand. Technol. **101**, 659–669 (1996).
[10]P. L. Carson, P. R. Fischella, and T. V. Oughton, "Ultrasonic power and intensities produced by diagnostic ultrasound equipment," Ultrasound

Med. Biol. **3**, 341–350 (1978).
[11]J. Lee and K. K. Shung, "Radiation forces exerted on arbitrarily located sphere by acoustic tweezer," J. Acoust. Soc. Am. **120**, 1084–1094 (2006).
[12]T. Hasegawa and K. Yosioka, "Acoustic-radiation force on a solid elastic sphere," J. Acoust. Soc. Am. **46**, 1139–1143 (1969).
[13]F. G. Mitri and S. Chen, "Theory of dynamic acoustic radiation force experienced by solid cylinders," Phys. Rev. E **71**, 016306 (2005).
[14]F. G. Mitri and M. Fatemi, "Dynamic acoustic radiation force acting on cylindrical shells: Theory and simulations," Ultrasonics **43**, 435–445 (2005).
[15]L. A. Crum, "Acoustic force on a liquid droplet in an acoustic stationary wave," J. Acoust. Soc. Am. **50**, 157–163 (1971).
[16]P. L. Marston, "Shape oscillation and static deformation of drops and bubbles driven by modulated radiation stresses—Theory," J. Acoust. Soc. Am. **67**, 15–26 (1980).
[17]S. Chen, M. Fatemi, and J. F. Greenleaf, "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889 (2002).
[18]W. F. Walker, F. J. Fernandez, and L. A. Negron, "A method of imaging viscoelastic parameters with acoustic radiation force," Phys. Med. Biol. **45**, 1437–1447 (2000).
[19]K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation force," J. Acoust. Soc. Am. **110**, 625–634 (2001).
[20]A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[21]M. Fatemi and J. F. Greenleaf, "Ultrasound-stimulated vibro-acoustic spectrography," Science **280**, 82–85 (1998).
[22]M. Fatemi and J. F. Greenleaf, "Vibro-acoustography: An imaging modality based on ultrasound-stimulated acoustic emission," Proc. Natl. Acad. Sci. U.S.A. **96**, 6603–6608 (1999).
[23]J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 396–409 (2004).
[24]S. A. McAleavey, M. Menon, and J. Orszulak, "Shear-modulus estimation by application of spatially-modulated impulsive acoustic radiation force," Ultrason. Imaging **29**, 87–104 (2007).
[25]S. Chen, M. Urban, C. Pislaru, R. Kinnick, Y. Zheng, A. Yao, and J. Greenleaf, "Shearwave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**, 55–62 (2009).
[26]S. Chen, G. T. Silva, R. R. Kinnick, J. F. Greenleaf, and M. Fatemi, "Measurement of dynamic and static radiation force on a sphere," Phys. Rev. E **71**, 056618 (2005).
[27]G. T. Silva, S. Chen, J. F. Greenleaf, and M. Fatemi, "Dynamic ultrasound radiation force in fluids," Phys. Rev. E **71**, 056617 (2005).
[28]M. W. Urban, G. T. Silva, M. Fatemi, and J. F. Greenleaf, "Multifrequency vibro-acoustography," IEEE Trans. Med. Imaging **25**, 1284–1295 (2006).
[29]G. T. Silva, M. W. Urban, and M. Fatemi, "Multifrequency radiation force of acoustic waves in fluids," Physica D **232**, 48–53 (2007).
[30]F. G. Mitri, J. F. Greenleaf, and M. Fatemi, "Chirp imaging vibro-acoustography for removing the ultrasound standing wave artifact," IEEE Trans. Med. Imaging **24**, 1249–1255 (2005).
[31]T. N. Erpelding, K. W. Hollman, and M. O'Donnell, "Bubble-based acoustic radiation force using chirp insonation to reduce standing wave effects," Ultrasound Med. Biol. **33**, 263–269 (2007).
[32]Y. Hu, D. Zhang, H. Zheng, and X. Gong, "Chirp excitation technique to enhance microbubble displacement induced by ultrasound radiation force," J. Acoust. Soc. Am. **125**, 1410–1415 (2008).
[33]A. B. Carlson, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication* (Irwin McGraw-Hill, Boston, MA, 1986).
[34]M. W. Urban and J. F. Greenleaf, "Harmonic pulsed excitation and motion detection of a vibrating reflective target," J. Acoust. Soc. Am. **123**, 519–533 (2008).
[35]R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, Boston, MA, 2000).
[36]M. W. Urban, R. R. Kinnick, and J. F. Greenleaf, "Measuring the phase of vibration of spheres in a viscoelastic medium as an image contrast modality," J. Acoust. Soc. Am. **118**, 3465–3472 (2005).
[37]H. L. Oestreicher, "Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics," J. Acoust. Soc. Am. **23**, 707–714 (1951).

Probing slow dynamics of consolidated granular multicomposite materials by diffuse acoustic wave spectroscopy

Nicolas Tremblay and Eric Larose[a)]
Laboratoire de Géophysique Interne et Tectonophysique, CNRS and Université J. Fourier, Grenoble BP 53 38041, France

Vincent Rossetto
Laboratoire de Physique et Modélisation des Milieux Condensés, CNRS and Université J. Fourier, Grenoble BP 166 38042, France

(Received 2 September 2009; revised 14 December 2009; accepted 29 December 2009)

The stiffness of a consolidated granular medium experiences a drop immediately after a moderate mechanical solicitation. Then the stiffness rises back toward its initial value, following a logarithmic time evolution called *slow dynamics*. In the literature, slow dynamics has been probed by macroscopic quantities averaged over the sample volume, for instance, by the resonant frequency of vibrational eigenmodes. This article presents a different approach based on diffuse acoustic wave spectroscopy, a technique that is directly sensitive to the details of the sample structure. The parameters of the dynamics are found to depend on the damage of the medium. Results confirm that slow dynamics is, at least in part, due to tiny structural rearrangements at the microscopic scale, such as inter-grain contacts. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3294553]

PACS number(s): 43.25.Ts, 43.35.Yb, 43.40.Le, 43.35.Zc [RLW] Pages: 1239–1243

I. INTRODUCTION: WHAT IS SLOW DYNAMICS?

Depending on the magnitude of the solicitation applied to it, a consolidated granular medium, such as sedimentary rocks or concrete, can asymptotically react either in the elastic (reversible) regime, or in the brittle regime, where irreversible fractures develop. A few years ago, several articles from Guyer and co-workers and Ten Cate and co-workers (Ten Cate and Shankland, 1996; Guyer *et al.*, 1998; Guyer and Johnson, 1999; Ten Cate *et al.*, 2000) reported on an intermediate regime, where they observed non-linear elasticity. This regime is observed after imposing a strain of moderate amplitude (10^{-6}) that does not generate any macroscopic damage. They called its time evolution as slow dynamics: it starts with a drop of the elastic modulus, followed by a logarithmic recovery of the sample stiffness toward its initial value after the strain is released. The $\log(t)$ recovery was found to be universal in granular solids of various composition (Ten Cate *et al.*, 2000; Lacouture *et al.*, 2003; Johnson and Sutin, 2005), humidity (Vakhnenko *et al.*, 2004), or level of damage (Van Den Abeele *et al.*, 2001). Most of the time, the evolution of the sample stiffness is monitored through the frequency of vibrational eigenmodes, using a technique named as non-linear elastic wave spectroscopy (Guyer and Johnson, 1999). More recently, Lobkis and Weaver (2009) proposed to monitor the stiffness of the material using the Larsen effect between two piezo-eletric transducers. The resonant frequency f_L observed with a feedback loop was measured with a relative precision of the order $\delta f_L/f_L \approx 10^{-6}$, and was found to satisfyingly monitor the rigidity of the material. This latter technique offers simultaneously an increased sensitivity and a temporal resolution of a few ms, thus allowing to study slow dynamics over very short time-scales.

In this article, we develop an alternative method based on broadband diffuse ultrasound. We take advantage of coda waves that can be observed in multicomposite materials, where multiple scattering effects (characterized by the scattering mean free time $t^\star$) dominate absorption effects (characteristic time τ_{abs}). Denoted by T, the central period of the ultrasound, and with Δd as the time between two ultrasonic records, the time-scales in our experiment follow the hierarchy

$$T < t^\star < \tau_{\mathrm{abs}} < \Delta d. \tag{1}$$

Similar to coda wave interferometry (CWI) (Poupinet *et al.*, 1984; Roberts *et al.*, 1992; Snieder *et al.*, 2002) and to diffuse acoustic wave spectroscopy (DAWS) (Pine *et al.*, 1988; Cowan *et al.*, 2002), we study weak changes in the diffuse waveforms in amplitude and phase. We compare waveforms recorded after a moderate solicitation to those recorded before it. The solicitation consists in the impact of a small steel ball falling onto the sample from a controlled height (1 m in general). The composition of the samples is displayed in Table I.

II. EXPERIMENTAL SETUP

In the first experiment, we use a cylindrical sample of 16 cm in diameter and 30 cm in height (sample 1). Two ultrasonic transducers are used: one as a source and one as a receiver (labeled as S and R, respectively, in Fig. 1); they are 10 cm apart. Note that the lateral part of the transducer is inactive and rigid, so that surface (Rayleigh) waves are neither excited nor recorded. In this sample, the scattering mean

[a)]Author to whom correspondence should be addressed. Electronic mail: eric.larose@ujf-grenoble.fr

TABLE I. Sample composition.

Composition (in weight)	Weight	Density
Cement: 17% Fine sand: 31% Gravel: 43% Water: 9%	12.5 kg	2.2 kg/l

free time is of the order of 12 μs, the transit time is of the order of 60 μs, and the absorption time is about 200 μs.

The lateral size of the active part of the transducers (0.7 mm) is much smaller than the wavelength (λ =5 mm–5 cm at working frequencies), which makes them very sensitive to multiply scattered waves. The transducers are glued onto the sample using a hot chemical glue (phenylsalicylic acid) that solidifies with cooling (below 43 °C). The ultrasonic experiment was performed several months after casting the samples. We send a 80 V broadband ultrasonic impulse with frequencies ranging from 50 to 500 kHz. The signal measured by the receiver is amplified, digitized, and stored in memory. A typical record $h_d(t)$ is shown in Fig. 2. The impulse response contains a hardly visible direct arrival, followed by a long lasting coda due to multiple scattered waves. The acquisition is repeated at various dates d (d=0 starts the experiment), over about 1 h. At date d_I, we drop a small metallic ball (m=30 g) from a height h=1.38 m onto the top face of the sample. The impact is about 15 cm from the source and receiver. The elastic energy released to the sample is $E_{\text{kin}}=mgh\approx 0.4$ J. Note that we observe no visible damage caused by this moderate impact.

III. DATA PROCESSING

From the set of ultrasonic data obtained during an experiment, we derive two quantities that depend on d. First, we measure the relative velocity change $\delta V_d/V$ within the material (Poupinet *et al.*, 1984; Snieder *et al.*, 2002; Lobkis and Weaver, 2003). Second, we measure the remnant decorrelation K_d of the waveforms after correcting for the effect of the relative velocity change (Lobkis and Weaver, 2003).

A. Relative velocity change $\delta V/V$ (CWI)

Let us describe the two steps of our data processing in details. To measure the relative velocity change, we compare the phase of the waveforms acquired at the date d to the phase of the initial record (d=0). Following Sens-Schönfelder and Wegler (2006), we interpolate and resample

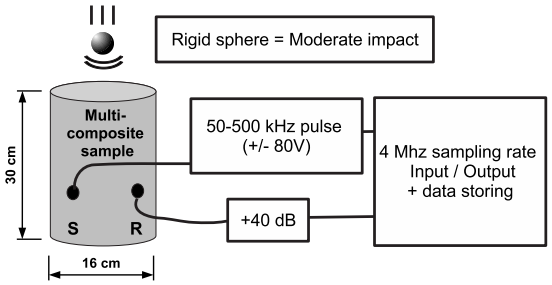


FIG. 1. Schematic view of the experimental setup for sample 1.

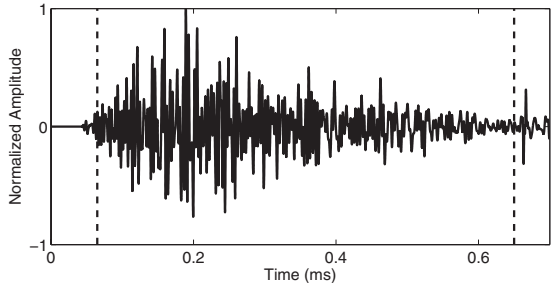


FIG. 2. Typical record collected through a concrete sample. The integration time-window is marked by the two dashed vertical lines at 0.05 and 0.65 ms. The initial impulse duration is about 10 μs.

$h_d(t)$ at times $[1+\epsilon]t$; this corresponds to stretching the time axis of the record by a factor of $1+\epsilon$. Then we evaluate the correlation coefficient between h_0 and h_d

$$X_d(\epsilon)=\frac{\int h_0(t)h_d([1+\epsilon]t)dt}{\sqrt{\int h_0^2(t)dt}\sqrt{\int h_d^2([1+\epsilon]t)dt}}, \tag{2}$$

where the integration is performed over a time-window larger than the period T. This time-window is marked by two vertical broken lines in Fig. 2. This calculation is repeated for various values of the parameter ϵ. The parameter ϵ maximizing X_d corresponds to the actual relative velocity change

$$\epsilon_{\max}=-\delta V_d/V. \tag{3}$$

The evolution of $\delta V_d/V$ as a function of d is displayed in Fig. 3 (top). Note that the ball impact occurs at date d_I =270 s after starting the experiment. Before the impact, there is no relative velocity change, indicating that the medium is at rest. Immediately after the impact, the velocity is noticeably decreased. Since the density of the material is not changed, and since diffuse waves are mainly in the trans-

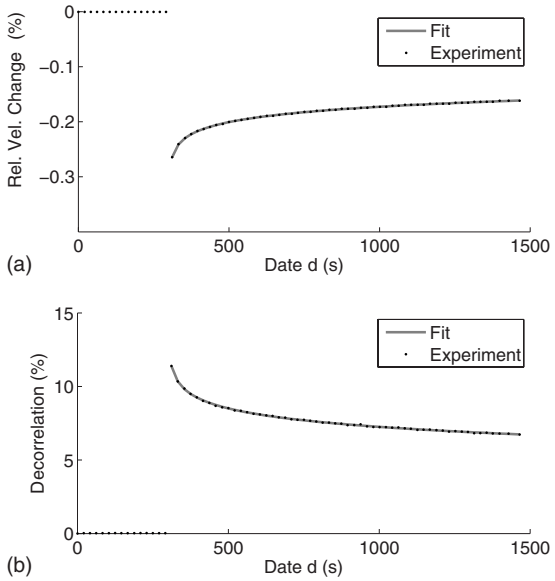


FIG. 3. Relative velocity change $\delta V/V$ (top) and decorrelation (bottom) versus the date d of the measurement. Impact is at d_I=270 s. Thermal effects are negligible. A logarithmic fit is added to each experimental curve.

verse (shear) polarization, this velocity drop is principally due to a drop of rigidity of the material. Then, the material stiffness rises back toward its initial value.

B. Remnant decorrelation *K* (DAWS)

In a second step, we use the dilated records to evaluate the remnant decorrelation coefficient K_d between the waveforms, which simply reads (Pine *et al.*, 1988; Cowan *et al.*, 2002)

$$K_d = 1 - X_d(\epsilon_{\max}). \tag{4}$$

The time evolution of K as a function of d is plotted in Fig. 3 (bottom). Again, the record is very stable (no change) before the impact. After the impact, the ultrasonic waveform differs significantly from the reference, then the decorrelation from the reference waveform decreases over several hundreds of seconds. The decorrelation K defined in Eq. (4) cannot be solely associated to a difference of relative velocity changes for different waves (for instance, between compressional and shear waves), contrary to the decorrelation studied by Lobkis and Weaver (2003). Also, the decorrelation cannot be attributed to changes in the bulk of the grains, which would require much more energy than the one released by the ball. Here, it corresponds to a feeble change in the scattering properties of the concrete sample, for instance, a small rearrangement of contacts between grains.

It is important to note that thermal changes are also well known to affect the average velocity within the sample (Lobkis and Weaver, 2003; Snieder *et al.*, 2002; Larose *et al.*, 2006), and consequently, the decorrelation. Relative velocity changes due to temperature variations are found to be negligible in our experiment, as can be noticed by the flat curve in Fig. 3, from $t=0$ to $t=270$ s. The decorrelation of the waveforms before the impact is actually near 0.1%, and is attributed to the noise generated by the electronic devices. We neglect it in the following. Note also that the changes in intensity of the diffuse waveforms (say, the quality factor of the coda) was observed to show some recovery (Ten Cate *et al.*, 2000), but this effect is not studied here and is beyond the scope of the present paper.

IV. THE PHYSICAL PARAMETERS OF SLOW DYNAMICS

The time evolutions of both $\delta V/V$ and K fit logarithmic laws (with d and d_I in seconds)

$$\delta V_d/V = A' - m' \log_{10}(d - d_I), \tag{5}$$

$$K_d = A - m \log_{10}(d - d_I). \tag{6}$$

Such a logarithmic evolution, valid if $d \ll 10^{A/m}$, is referred to as slow dynamics. A and m are the fit parameters, where A is the extrapolated decorrelation 1 s after the impact, and m is the slope of the logarithmic decay. $10^{A/m}$ represents the recovery time (in seconds), which is the time the system needs to recover its initial state. In general, this recovery time is of the order of 10^4–10^5 s. Parameters A and m were found to depend on the experimental apparatus, but Eq. (6) was found to describe the dynamics in all cases. Ten Cate *et al.* (2000) mentioned a break away from the log recovery after 1000 s. In our experiments, we should certainly expect an erosion of the log dynamics at some time. In particular, we must consider the possibility that part of the change in our experiments might be irreversible, and that the recovery might saturate. This could be ascribed to irreversible (but invisible) damages caused by the impact. As we could not monitor the sample longer than 1000 s, we cannot address this issue properly.

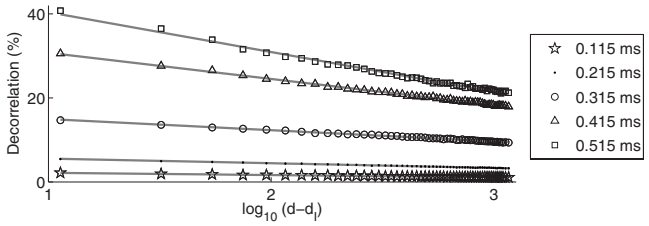


FIG. 4. Decorrelation versus the date of the measurement (in log scale) for various time-windows in the coda. The central time of each window is displayed in the box. The recovery time $10^{A/m}$ is found to be constant in the whole record.

A. Slow dynamics versus time in the coda

Let us now study the parameters A and m in more detail. As a first test, we evaluate these parameters at different times t in the coda. To that end, we split the initial time window (see Fig. 2) into five consecutive time-windows with length of 0.12 ms. The decorrelation K_d for these five windows is displayed in Fig. 4. A and m were found to increase with the time in the coda, confirming that late diffuse waves are much more sensitive to weak perturbation than early arrivals. Interestingly enough, the recovery time $10^{A/m}$ is constant within the whole record, which confirms that all time-windows (thus all wave paths) undergo the same slow dynamics phenomenon.

B. Slow dynamics versus frequency

To study the frequency dependence of slow dynamics, we filter the whole record into different frequency bands; we observe that A/m is constant in the 50–500 kHz frequency range. That the same dynamics was observed at different wavelength might suggest that there is no characteristic length in the rearrangement of the material structure. This confirms the assumption commonly made by previous authors (Ten Cate *et al.*, 2000) that the changes occur at various scales, following the probable continuous distribution of energy barriers.

C. Spatial dependence of slow dynamics

In the initial works on slow dynamics (Ten Cate and Shankland, 1996; Guyer *et al.*, 1998), the medium was subject to a global excitation, and resulted in a global change in the sample stiffness. The excitation presented here is a local impact (as in (Lobkis and Weaver, 2009). Determining whether the change induced by the ball impact is global (induced by the shock wave), or local is of some importance. To that end, we deploy a linear array of four sensors 10 cm

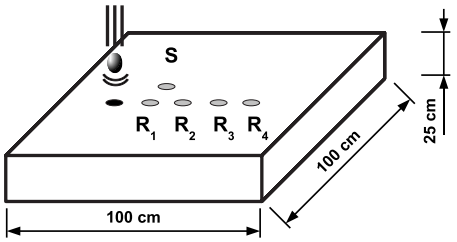


FIG. 5. Experimental setup for sample 2. Four receivers (R_{1-4}) are aligned with the impact location. The source S is 10 cm off the impact-receiver axis.

away from the impact location and from the source transducer. Because the initial sample is not large enough, we use a different sample (2) of similar composition but whose dimensions are now $1\times1\times0.25$ m^3 (see Fig. 5). The array has a spatial pitch of about 10 cm. The decorrelation K is found to strongly depend on the impact-receiver distance (Fig. 6). We therefore confirm the assumption issued by Lobkis and Weaver (2009) that the changes induced by the ball impact are located in the vicinity of the impact, and the seismic (shock) wave emitted by the impact has negligible effect on the material stiffness at large distances.

V. TENTATIVE APPLICATION TO ON-SITE DAMAGE ESTIMATION

Last, we address the issue of damage estimation through DAWS. Consider two blocks of concrete of similar constitution (see Table I) and of the same geometry (see sample 1); the first one is intact and the other one is damaged after undergoing a 30 MPa load. Both samples are tested using the same experimental protocol, and data are processed in the same time window and frequency range. Resulting decorrelations K_d are plotted in Fig. 7. Similar to Van Den Abeele *et al.* (2001), we observe that damage increases considerably the amplitude of change in the material, and thus, the absolute decorrelation A together with the speed of recovery m. Provided that the experimental protocol is rigorously similar from one site to another, this latter result suggests a possible route for on-site concrete damage assessment.

VI. CONCLUSION

In this article, we studied the slow dynamics behavior of concrete after a ball impact of moderate energy. We observed

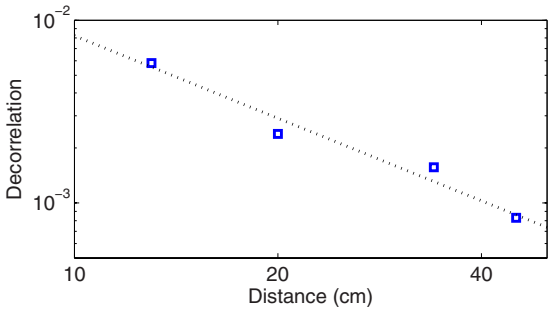


FIG. 6. (Color online) Decorrelation of ultrasonic waveforms obtained a few ms right after the impact in log-log scale, at different distances from the impact. Squares represent the experimental data, while dots represent a power law with the exponent −1.5.

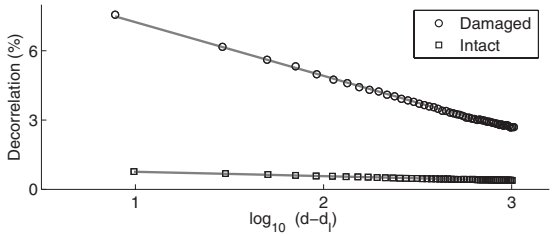


FIG. 7. Slow dynamics observed in undamaged and damaged concretes.

a drop of the acoustic velocity (probed by CWI, Poupinet *et al.*, 1984) and a decorrelation of the diffuse waveforms (probed by DAWS). Then the material recovers over characteristic times ranging from 100 to several 1000 s. In our experiments, the $\log(t)$ recovery was observed at all time in the coda, and at different frequencies; this does not allow to exhibit any characteristic length, time, or energy. This feature is in agreement with a fractal structure of concrete. Slow dynamics is thus interpreted as a series of jumps of energy barriers having a continuous distribution. In principle, slow dynamics could be either due to a homogeneous change in mechanical characteristics of all constituents, or to local rearrangements of inter-grain contacts. As we observe a noticeable decorrelation of the diffuse waveforms that cannot be solely ascribed to a homogeneous change in the material stiffness, we defend the second interpretation; slow dynamics is due to a collection of local changes in inter-grain contacts, which simultaneously result in changes in scattering properties and in a decrease in the effective wave velocity. The rearrangements of inter-grain contacts in our experiments are somehow analogous to those observed in solid friction (Lomnitz, 1962; Pandit and Savage, 1973; Karner and Marone, 1998). As concrete is an example of consolidated multicomposite granular material, we also believe that our laboratory experiment shares analogy with the time evolution of shallow sedimentary geomaterials after an earthquake (Sawazaki *et al.*, 2006; Brenguier *et al.*, 2008). In our experiments, slow dynamics parameters A and m where found to be very sensitive to experimental apparatus, particularly to the impact energy and position. While keeping these features perfectly constant and the experiment perfectly reproducible, we observed that the slow dynamics parameters were also found to strongly depend on damage, which suggest that they could characterize the aging of concrete structures. Since the experimental apparatus is light and easy to handle, it could have some practical interest in structural health monitoring, especially in civil engineering, or to perform down hole geomaterial characterization, provided that a reproducible and controlled impact could be delivered to the medium at depth.

ACKNOWLEDGMENTS

The authors would like to thank R. L. Weaver and P. A. Johnson for attracting their attention on slow dynamics. This work was funded by the ANR SISDIF Grant No. 08-JCJC-0066.

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., and Larose, E. (**2008**). "Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations," Science

321, 1478–1481.
Cowan, M. L., Jones, I. P., Page, J. H., and Weitz, D. A. (**2002**). "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**, 066605.
Guyer, R. A., and Johnson, P. A. (**1999**). "Non-linear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**(4), 30–36.
Guyer, R. A., McCall, K. R., and Abeele, K. V. D. (**1998**). "Slow elastic dynamics in a resonant bar of rock," Geophys. Res. Lett. **25**, 1585–1588.
Johnson, P., and Sutin, A. (**2005**). "Slow dynamics and anomalous fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**, 124–130.
Karner, S. L., and Marone, C. (**1998**). "The effect of shear load healing in simulated fault gouge," Geophys. Res. Lett. **25**, 4561–4564.
Lacouture, J., Johnson, P., and Cohen-Tenoudji, F. (**2003**). "Study of critical behavior in concrete during curing by application of dynamic linear and nonlinear means," J. Acoust. Soc. Am. **113**, 1325–1332.
Larose, E., de Rosny, J., Margerin, L., Anache, D., Gouedard, P., Campillo, M., and VanTiggelen, B. (**2006**). "Observation of multiple scattering of kHz vibrations in a concrete structure and application to weak changes monitoring," Phys. Rev. E **73**, 016609.
Lobkis, O. I., and Weaver, R. L. (**2003**). "Coda-wave interferometry in finite solids: Recovery of P-to-S conversion rates in an elastodynamic billiard," Phys. Rev. Lett. **90**, 254302.
Lobkis, O. I., and Weaver, R. L. (**2009**). "On the Larsen effect to monitor small fast changes in materials," J. Acoust. Soc. Am. **125**, 1894–1905.
Lomnitz, C. (**1962**). "Application of the logarithmic creep low to stress wave attenuation in the solid earth," J. Geophys. Res. **67**, 365–368.
Pandit, B. I., and Savage, J. C. (**1973**). "An experimental test of Lomnitz's theory of internal friction in rocks," J. Geophys. Res. **78**, 6097–6099.
Pine, D. J., Weitz, D. A., Chaikin, P. M., and Herbolzheimer, E. (**1988**). "Diffusing-wave spectroscopy," Phys. Rev. Lett. **60**, 1134–1137.
Poupinet, G., Ellsworth, W. L., and Frechet, J. (**1984**). "Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California," J. Geophys. Res. **89**, 5719–5731.
Roberts, P. M., Phillips, W. S., and Fehler, M. (**1992**). "Development of the active doublet method for measuring small velocity and attenuation changes in solids," J. Acoust. Soc. Am. **91**, 3291–3302.
Sawazaki, K., Sato, H., Nakahara, H., and Nishimura, T. (**2006**). "Temporal change in site response caused by earthquake strong motion as revealed from coda spectral ratio measurement," Geophys. Res. Lett. **33**, L21303.
Sens-Schönfelder, C., and Wegler, U. C. (**2006**). "Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano, Indonesia," Geophys. Res. Lett. **33**, L21302.
Snieder, R., Grêt, A., Douma, H., and Scales, J. (**2002**). "Coda wave interferometry for estimating nonlinear behavior in seismic velocity," Science **295**, 2253–2255.
Ten Cate, J. A., and Shankland, T. J. (**1996**). "Slow dynamics in the nonlinear elastic response of Berea sandstone," Geophys. Res. Lett. **23**, 3019–3022.
Ten Cate, J. A., Smith, E., and Guyer, R. A. (**2000**). "Universal slow dynamics in granular solids," Phys. Rev. Lett. **85**, 1020–1023.
Vakhnenko, O. O., Vakhnenko, V. O., Shankland, T. J., and TenCate, J. A. (**2004**). "Strain induced kinetics of inter-grain defects as the mechanism of slow dynamics in the nonlinear resonant response of humid sandstone samples," Phys. Rev. E **70**, 015602.
Van Den Abeele, K., Sutin, A., Carmeliet, J., and Johnson, P. (**2001**). "Micro-damage diagnostics using nonlinear elastic wave spectroscopy (NEWS)," NDT Int. **34**, 239–248.

Detecting blast-induced infrasound in wind noise

Wheeler B. Howard[a)]
University of Mississippi National Center for Physical Acoustics, 1 Coliseum Drive, University, Mississippi 38677

Kevin L. Dillion
Miltec Research and Technology, 9 Industrial Park Drive, Oxford, Mississippi 38655

F. Douglas Shields
University of Mississippi National Center for Physical Acoustics, 1 Coliseum Drive, University, Mississippi 38677

(Received 25 June 2009; revised 18 December 2009; accepted 20 December 2009)

Current efforts seek to monitor and investigate such naturally occurring events as volcanic eruptions, hurricanes, bolides entering the atmosphere, earthquakes, and tsunamis by the infrasound they generate. Often, detection of the infrasound signal is limited by the masking effect of wind noise. This paper describes the use of a distributed array to detect infrasound signals from four atmospheric detonations at White Sands Missile Range in New Mexico, USA in 2006. Three of the blasts occurred during times of low wind noise and were easily observed with array processing techniques. One blast was obscured by high wind conditions. The results of signal processing are presented that allowed localization of the blast-induced signals in the presence of wind noise in the array response. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3291683]

PACS number(s): 43.28.Dm, 43.28.Mw, 43.60.Gk [VEO] Pages: 1244–1250

I. INTRODUCTION

Infrasound signals are classified as acoustic waves with frequencies below 20 Hz. These low frequency waves can propagate hundreds to thousands of kilometers depending on atmospheric conditions. This facet of long-range propagation makes infrasound useful in remotely detecting different levels of naturally occurring and man made events. A snapshot of some current infrasound research fields includes volcanic eruptions,[1–3] conventional and nuclear blast monitoring,[4,5] earthquake,[6,7] tsunamis[6,8] surf-shore interactions,[9–12] atmospheric phenomena,[13,14] and bolide/meteor entry into Earth's atmosphere.[15–17]

Applications for infrasound in each of these research fields suffer from a common problem; the background acoustic noise floor is a function of the wind velocity.[18,19] Winds of only a few m/s can raise the noise floor significantly, decreasing the signal-to-noise ratio, masking infrasound signals from events of interest. A technique commonly employed to separate infrasonic signals from wind noise involves spatially averaging the atmospheric pressure variations. While infrasonic signals are coherent over hundreds of meters,[20] pressure variations due to the wind are coherent over a few meters; thus, spatial averaging reduces the wind noise but not the coherent infrasound. The averaging is usually accomplished mechanically through the use of pipe or hose arrays.

Typical International Monitoring System (IMS) pipe arrays consist of multiple ports open to the atmosphere deployed in a rosette design with an array diameter usually between 30 and 70 m.[21] The pressure measured at a highly sensitive microbarometer has passed through a central manifold equidistant from all of the ports. Thus, the measured pressure is a spatial average of the pressure at the ports. Because the coherent length of wind is small with respect to the array aperture, local pressure variations due to wind advection are summed incoherently. Acoustic signals with wavelengths that are long relative to the array aperture constructively add at the summing manifold. Spatial averaging provides an amplitude signal-to-noise gain of the square root of the number of openings to the atmosphere. This type of array is currently deployed at Piñon Flat, CA.

Portable infrasound arrays and the IMS array employed at Manoa, HI are examples of the porous hose array. Here, multiple hoses are connected to a summing manifold which leads to a centrally located microbarometer. Spatial averaging with porous hose arrays is accomplished along the length of the porous hoses[22] instead of at the discreet ports of pipe arrays. It has been suggested that these arrays may provide better wind noise reduction than pipe arrays.[23] Mobile arrays of this type are deployed in the southeastern United States to investigate hurricanes in the Gulf of Mexico. Although porous hose arrays are easily deployed and mobile, their performance depends on their manufacture and can be compromised by degradation of the porous hose material by solar radiation.

An alternate method of spatial averaging is accomplished through the use of a distributed array.[20] Here, multiple microphone elements are deployed over an area. This type of array provides multiple pressure measurements over the array aperture compared to a single measurement from the pipe and hose arrays. The performance of this type of array is dependent on multiple factors: microphone sensitiv-

[a)]Author to whom correspondence should be addressed. Electronic mail: wbhoward@olemiss.edu

 0001-4966/2010/127(3)/1244/7/$25.00

ity, number of sensors, sensor spacing, and array aperture. When the sensor spacing is such that the wind noise is incoherent between sensors, averaging the data from all the sensors at each recorded time step decreases the pressure amplitude of the wind noise by a factor of the square root of the number of sensors. As well as being mobile, an advantage of distributed arrays is that their data are amenable to various signal processing techniques.

II. DATA ANALYSIS: F-STATISTIC AND THE COEFFICIENT OF VARIATION

There is a need to scan large quantities of digitized data to identify events of interest. Therefore, an active area of interest is the development of detectors to locate events of interest. This enables the analyst to look at a small subset of the data when events of interest may have occurred. In the infrasound community the analysis of variance and Fisher's F-statistic are often used to indicate the presence of a coherent infrasound signal occurring on multiple sensors.[24] This test is a hypothesis test to determine if samples can be viewed as resulting from the same process. An event is deemed statistically important if the value of the F-statistic is larger than the associated critical value, calculated from the F cumulative distribution function.

Calculations of the F-statistic are found in the literature;[25–27] thus only a brief review will be provided here. Assume a data set from multiple sensors in tabular form. Further assume that each row represents the pressure recorded by all sensors at each time step so that the data from a singular sensor as a function of time are contained within a column. For T time steps and S sensors this data table would represent a $T \times S$ matrix. The value of the grand mean of all the samples is

$$\bar{X} = \frac{1}{TS} \sum_{j=1}^{T} \sum_{k=1}^{S} X_{jk}, \tag{1}$$

while the mean of each row would be

$$\bar{X}_j = \frac{1}{S} \sum_{k=1}^{S} X_{jk}. \tag{2}$$

This quantity is the mean pressure value over all sensors at the jth instant in time. The total variation is the sum of the squares of the deviations of each measurement from the grand mean,

$$V = \sum_{j=1}^{T} \sum_{k=1}^{S} (X_{jk} - \bar{X})^2. \tag{3}$$

After some algebraic manipulation, the total variation can be broken up into the sum of V_w and V_B, where

$$V_W = \sum_{j=1}^{T} \sum_{k=1}^{S} (X_{jk} - \bar{X}_j)^2 \tag{4}$$

and

$$V_B = S \sum_{j=1}^{T} (\bar{X}_j - \bar{X})^2. \tag{5}$$

The F-statistic, calculated from these quantities, is given by

$$F = \frac{V_B/(T-1)}{V_W/(T(S-1))}. \tag{6}$$

If the hypothesis that these data result from the same process is true, then the calculated F-value will be unity. This value will increase as the difference between the pressures at the sensors increases.

A second detector is based on the reciprocal of the coefficient of variation[28] (RCOV) ratio. Again assuming data from sensors in tabular form with the output from a single sensor being contained within a column, the mean value of the pressure over the array at the jth time instant is

$$\bar{X}_j = \frac{1}{S} \sum_{k=1}^{S} X_{jk}. \tag{7}$$

The standard deviation σ about the mean array pressure at each time step is

$$\sigma = \left[\frac{1}{S} \sum_{k=1}^{S} (X_{jk} - \bar{X}_j)^2 \right]^{1/2}. \tag{8}$$

The magnitude of the RCOV, calculated as the ratio of the mean to the standard deviation, is

$$\mathrm{RCOV} = \frac{|\bar{X}_j|}{\sigma}. \tag{9}$$

If the separation of the sensors in the array is larger than the coherent length of the wind, then spatially averaging the output from the elements reduces the wind noise amplitude by a factor of $\sqrt{S}$. It is thought that the RCOV will be most effective when the wavelength of the signal is much longer than the aperture of the array. For such acoustic signals, the associated pressure signal varies little over the distributed array elements. Thus, the standard deviation of the sensor signals is small and RCOV is large. For signals with wavelengths comparable to or shorter than the array aperture, the standard deviation resulting from the pressure signal would be large and the magnitude of the RCOV would be small. These two detectors were utilized in examining data from atmospheric explosions.

III. EXPERIMENTAL SETUP

A. Field setup and data collection

The data reported here were collected during two experiments conducted near the White Sands Missile Range (WSMR) in March (WSMR2) and July (WSMR3) of 2006.[29] For each experiment two rockets were launched, one occurring around 2 a.m. and the next around 6 a.m. Explosive charges for each launch were approximately 30 kg trinitrotoluene (TNT) equivalent. These charges were detonated near altitudes of 35 km (WSMR2) and 49 km (WSMR3).

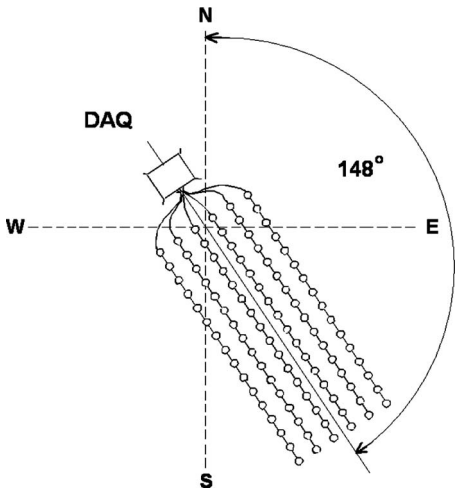


FIG. 1. Schematic of the distributed array layout and orientation for WSMR3.

The along the ground distance from the source to the distributed array was approximately 117 km (WSMR2) and 110 km (WSMR3).

Data for both experiments were collected at a site near Socorro, NM. For these tests a 93-sensor distributed array was deployed. Each array element was composed of piezounimorph elements housed in an engineered polyvinyl chloride (PVC) capsule.[20] These sensors exhibit a flat frequency response between 0.1 and 100 Hz relative to a 1/2 in. Brüel & Kjær 4193L-004 microphone. Signals output from each of the 93 sensors were digitized at 200 Hz by three National Instruments 6031E analog to digital converters. To reduce cross talk, a channel shorted to ground was read between each sensor signal reading.

The array setup for the two experiments was similar. For both tests the sensors were laid out in a rectangular aperture, as shown in Fig. 1. The array consisted of six rows with 16 sensors in each row. Spacing between rows was 2 m and sensors in each row were separated by 2 m, giving an array aperture of 30 m $\times$ 10 m. Shorted channels were substituted for sensors 1, 33, and 65. For the WSMR2 experiment, the rows were oriented along $144 \pm 1°$, relative to local magnetic North (0°). During the WSMR3 experiment the rows were oriented along $148 \pm 1°$, relative to magnetic local North. These orientations pointed the long arm of the array in the approximate direction of the explosion at detonation.

Wind speeds during the two WSMR2 tests were measured with a tri-axial Campbell Instrument Co. CSAT3 ultrasonic anemometer located near sensor 1. This anemometer, deployed 10 ft above the ground, provided simultaneous wind velocity data during the pressure measurements. Average wind speeds during the 2 a.m. and 6 a.m. detonations were 0.45 and 0.65 m/s, respectively. For the WSMR3 experiments, wind speeds were measured at 10 Hz by a tri-axial Vaisala ultrasonic anemometer located about 640 m away from the array. Figure 2 shows the measured wind speed for 15 min windows encompassing the 2 a.m. and 6 a.m. events. During the 2 a.m. event, the average wind speed was 5.0 m/s with gusts of almost 8.8 m/s while wind speeds for the 6 a.m. shot averaged 1.5 m/s with gusts of up to 2.7 m/s. Thus, the 2 a.m. wind speeds are three to four times greater than wind speeds from the 6 a.m. shot.

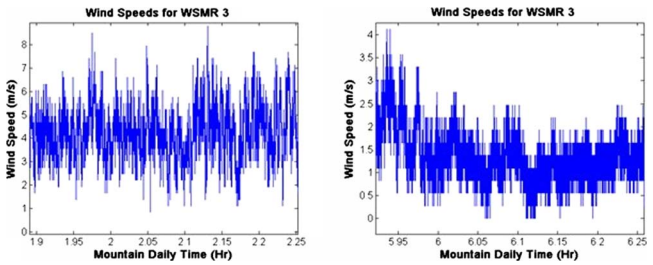


FIG. 2. (Color online) Measured wind speeds during the WSMR3 2 a.m. (left) and 6 a.m. (right) detonations.

B. Data processing

Data from each of the WSMR experiments were digitally filtered between 0.1 and 99 Hz with a fourth order Butterworth bandpass filter. The Butterworth filter was chosen as it has a maximally flat response and does not introduce ripple into the passband.[27] The mean pressure of the 93 sensors at each instant of recorded time was calculated. Data input to the RCOV and F-statistic calculations was further bandpass filtered between 1 and 10 Hz with the same fourth order digital Butterworth filter. The RCOV was calculated at each instant of recorded time over the entire file. A second program was employed to calculate the F-statistic value and associated critical value. This calculation was performed over 200 point bins, yielding a value of the F-statistic for each bin. The lag times between sensors, calculated via the cross correlation, were used to time shift the data before calculation of the F-statistic.

IV. EXPERIMENTAL DATA

A. WSMR2: Filtered and processed 2 a.m. shot

The digitally filtered (0.1–99 Hz) pressure data from sensor 39, located near the center of the array, are shown in the top graph of Fig. 3. For this shot the mean wind speed was 0.45 m/s. Several possible detonation arrivals are observed in the pressure versus time graph from the single sensor. A single infrasonic arrival is clearly indicated by the mean pressure from all 93 sensors at each recorded time (the bottom graph of Fig. 3). The insets in the top and bottom graphs of Fig. 3 show the pressure data for a 5 s window which included the blast arrival. Spatial averaging, without any time shifting to align the data from the sensors, reduced the wind noise pressure fluctuations observed with the single sensor response. This facilitated identification of the infrasonic arrival about 761 s into the WSMR2 2 a.m. data file. The mean peak-to-peak acoustic pressure over the array associated with the blast was about 0.6 Pa.

The raw pressure data from each sensor were then filtered between 1 and 10 Hz with a fourth order Butterworth filter. These filtered data served as input to programs that calculated the RCOV magnitude and the F-statistic value. In this analysis, the F-statistic value was computed over every 200 points (1 s). The pressure signal associated with the blast arrival is readily observed in both the RCOV magnitude (top

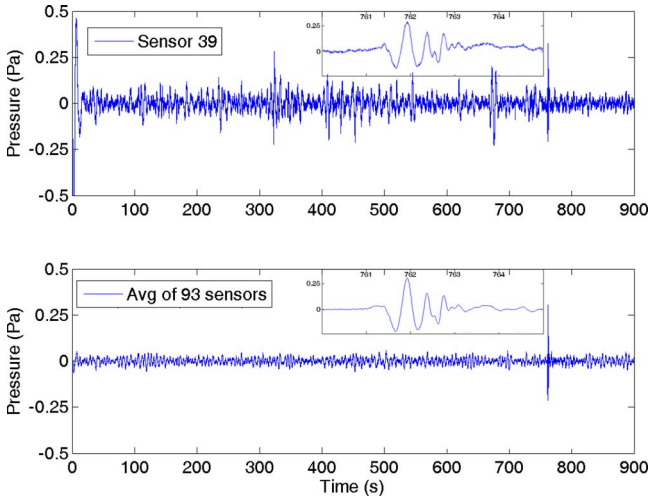


FIG. 3. (Color online) (Top) Pressure data from a single sensor near the middle of the array. (Bottom) The average pressure over the entire array during the 2 a.m. WSMR2 shot. The insets show the pressure data for the respective graphs over a 5 s window starting at 760 s.

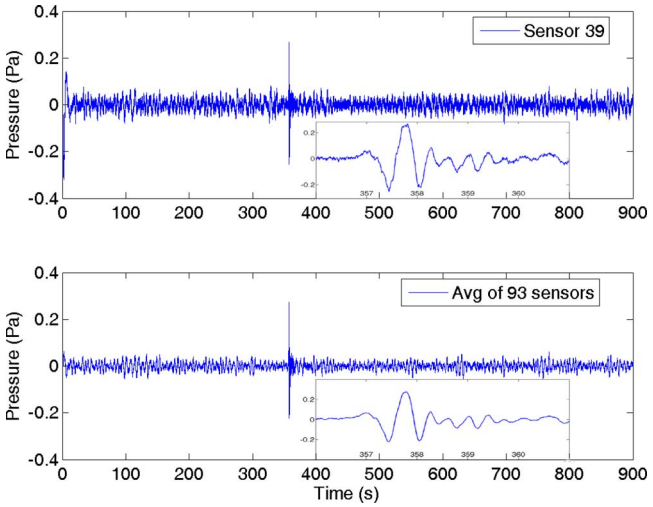


FIG. 5. (Color online) (Top) Pressure data from a single sensor near the middle of the array. (Bottom) The average pressure over the entire array during the 6 a.m. WSMR2 shot. The insets show the pressure data for the respective graphs over a 5 s window starting at 356 s.

graph Fig. 4) and F-statistic (bottom graph Fig. 4). The temporal location of the peak in both the RCOV and F-statistic graphs agrees with the arrival as observed in the average of the pressure data (bottom graph Fig. 3). Both the RCOV and F-statistic metrics indicate good agreement in localizing in time the blast event. The critical value for the 99.9% confidence level is 1.5201.

B. WSMR2: Filtered and processed 6 a.m. data

The 0.1–99 Hz digitally filtered WSMR2 6 a.m. pressure data are shown in Fig. 5 for both sensor 39 and the array mean. During this shot the mean wind speed was 0.65 m/s. In this figure the wind noise at sensor 39 does not mask the pressure associated with the blast arrival in the pressure versus time curve. The insets in the top and bottom graphs of Fig. 5 show the pressure data for a 5 s window which included the blast arrival. Calculation of the spatial average of the pressure further decreases the incoherent wind noise and indicates the arrival of the blast signal. For both the 2 a.m. and 6 a.m. WSMR2 measurements, the mean acoustic pressure over the array was about 0.6 Pa peak to peak.

After filtering the raw pressure data from each sensor between 1 and 10 Hz, the RCOV and F-statistic from these data were calculated. These results are observed in Fig. 6. The pressure data and temporal location of the peak in the RCOV and F-statistic graphs both indicate the blast arrival at 358 s in the recorded data. These both agree with the mean array response shown in the pressure curves of Fig. 5. Here also, the RCOV and F-statistic results indicate good agreement for detecting the blast event in low wind conditions.

C. WSMR3: Filtered and processed 6 a.m. shot

Pressure data collected from sensor 39, located near array center, and the mean of all 93 sensors for the 6 a.m. launch are shown in the top and bottom graphs of Fig. 7. These data have been bandpass filtered between 0.1 and 99 Hz. As seen from the top graph, the data from sensor 39 show instances of relatively low and high pressures. Some of these pressure events are comparable to the expected 0.6 Pa peak-to-peak amplitudes observed in both the WMSR2 tests. Comparison of the pressure data from sensor 39 to the wind

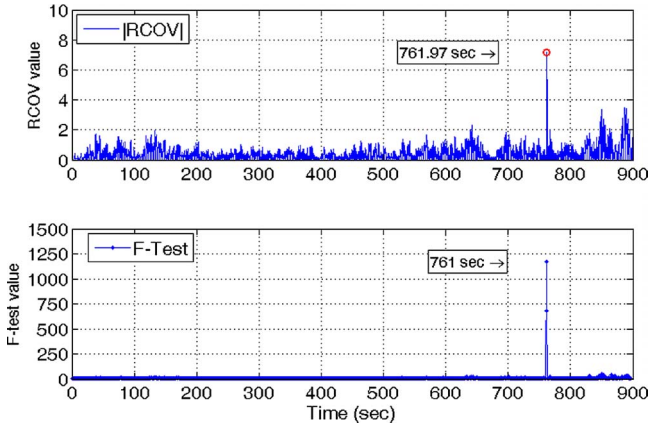


FIG. 4. (Color online) Results of the RCOV (top) and F-test (bottom) analyses on the 2 a.m. WSMR2 data.

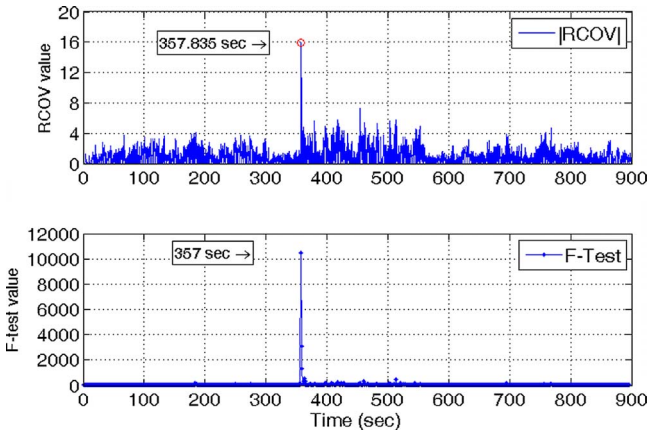


FIG. 6. (Color online) Results of the RCOV (top) and F-test (bottom) analyses on the 6 a.m. WSMR2 data.

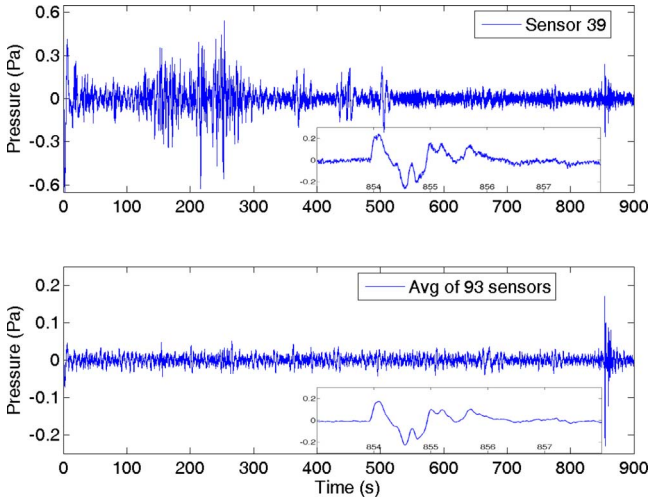


FIG. 7. (Color online) (Top) Pressure data from a single sensor near the middle of the array. (Bottom) The average pressure over the entire array during the 6 a.m. WSMR3 shot. The insets show the pressure data for the respective graphs over a 5 s window starting at 853 s.

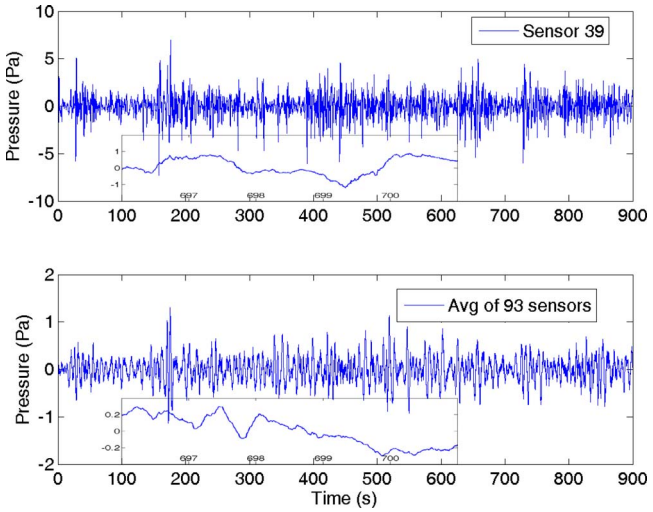


FIG. 9. (Color online) (Top) Pressure data from a single sensor near the middle of the array. (Bottom) The average pressure over the entire array during the 2 a.m. WSMR3 shot. The insets show the pressure data for the respective graphs over a 5 s window starting at 696 s.

speed data for this test (right graph, Fig. 2) indicates that larger pressure fluctuations result from higher wind speeds. The bottom graph shows the pressure calculated from averaging the data from all 93 sensors at each recorded time step. In this bottom graph, the incoherent wind noise reduced by the averaging process reveals the arrival of the infrasound signal around 855 s into the file. Although this arrival is also observed in the data from the single sensor, identifying this event in the pressure signal is problematic since the pressure amplitude resulting from the wind noise is comparable to the infrasound signal. The pressure associated with the blast is about a 0.5 Pa peak-to-peak signal in these pressure curves. This agrees with the data from the two WSMR2 tests. The insets in the top and bottom graphs of Fig. 7 show a close-up view of the pressure data for a 5 s window inclusive of the blast arrival.

The raw pressure data from each sensor were further digitally filtered between 1 and 10 Hz. Calculations of the RCOV magnitude and F-statistic values, shown in Fig. 8, provide confirmation that the signal observed around 855 s is indeed the blast-induced infrasonic signal of interest. Comparison between the data from the single sensor (top graph of Fig. 7) and the RCOV magnitude (top graph of Fig. 8) indicates that this measure shows low values during high wind conditions (right graph of Fig. 2) and higher values during low wind noise/coherent signal events.

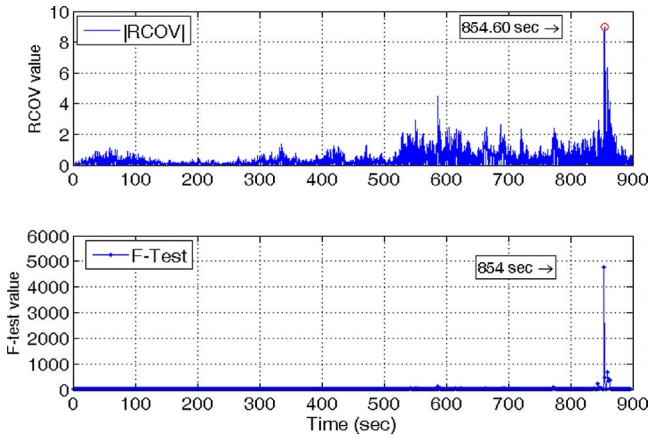


FIG. 8. (Color online) Results of the RCOV (top) and F-test (bottom) analyses on the 6 a.m. WSMR3 data.

D. WSMR3: Filtered and processed 2 a.m. data

The filtered data from sensor 39 and the average of all 93 sensors for the 2 a.m. shot are shown in Fig. 9. Pressure data from sensor 39 are nearly an order of magnitude larger than the data from the same sensor taken 4 h later. This is a result of the higher wind velocities during this test compared to any of the previous tests. In the pressure response from sensor 39, an infrasound signal of about 0.5 Pa, similar to that observed in the 6 a.m. data, would be masked by the wind noise. Computing the spatial average of the data from all 93 sensors (bottom graph of Fig. 9) does not identify the blast signal. Although the averaging reduced the incoherent wind noise, the pressure fluctuations in the spatial average are comparable or larger than the expected 0.5 Pa peak-to-peak signal. The blast arrival is also not readily observed in the expanded pressure inset plots in the top and bottom graphs of Fig. 9. Multiple bandpass filter values were employed to determine if better delineation of the signal arrival in the time-pressure plots could be achieved. None of these investigations proved fruitful in localizing the signal arrival in time.

For these high wind noise data, calculating the RCOV magnitude proved especially beneficial. After filtering the data between 1 and 10 Hz, the RCOV magnitude and F-statistic values were calculated. Results of this analysis are shown in Fig. 10. An event at about 698 s into the RCOV calculation stands out. Further investigation of this event shows it to be consistent with the arrival of the blast infrasound. Calculation of the F-statistic with a 200 point (1 s) window did not provide identification of the infrasonic blast event.

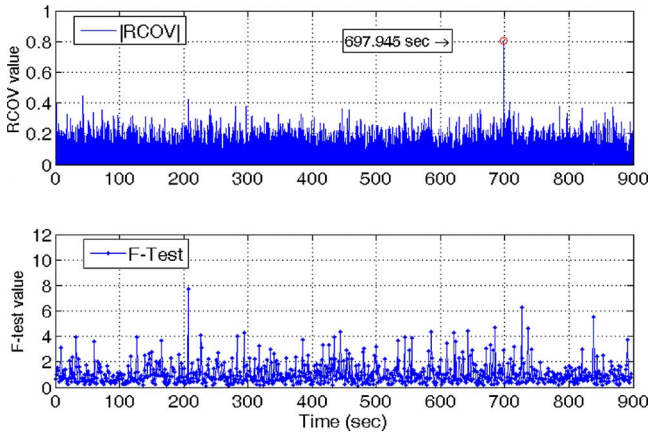


FIG. 10. (Color online) Results of the RCOV (top) and F-test (bottom) analyses on the 2 a.m. WSMR3 data.

V. DISCUSSION AND CONCLUSION

Both the F-statistic and magnitude of the RCOV provided clear indication of coherent sound propagation across the array in low wind conditions. Analysis of data taken when the mean wind speed about 5 m/s indicates that the RCOV technique presented above permits better temporal localization of the infrasound event. This agrees with the comment[24] that the F-statistic does not perform well for low signal-to-noise ratio events or when multiple signals are present in the data.

Smaller bin sizes were investigated to ascertain if newly calculated F-statistic values for the WSMR3 2 a.m. data would indicate the infrasonic arrival. Altering the F-statistic algorithm to calculate the F-value over a 20 point (0.1 s) window did provide indication of the blast event. This result is shown in Fig. 11. The value of the F-statistic corresponding to the infrasound arrival is greater than the 99.9% critical value obtained from Fisher's F probability function.

A significant percentage of the calculated F-statistic values is larger than the associated 99.9% F-critical values. Decreasing the size of the bins for analysis of the F-statistic did not show any appreciable difference in the percentage of calculations above the F-critical value threshold, except for the WSMR3 2 a.m. data. As an example, the F-statistic values for the WSMR2 2 a.m. data were greater than the critical value in 99%, 93%, and 62% of the bins when the bin sizes were 1000, 200, and 20 points, respectively.

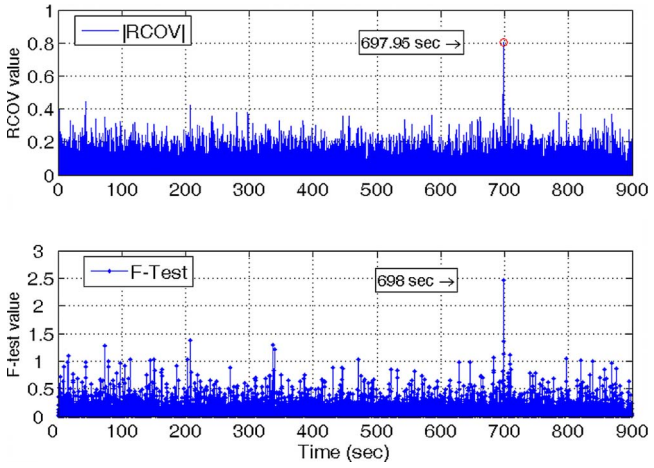


FIG. 11. (Color online) Results of the RCOV (top) and F-test (bottom) analyses on the 6 a.m. WSMR3 data with a 0.1 s (20 point) data window.

Pressure fluctuations due to wind noise present problems in the collection and interpretation of infrasound signals. A distributed array has been used to mitigate the influence of wind noise on infrasound signals of interest. Signals from four high altitude explosions in markedly different wind conditions have been collected. In each case the peak-to-peak pressure amplitude of the signals was between 0.5 and 0.6 Pa. In the cases of low wind velocity it was not always possible to clearly identify the arrival time of the blast-induced infrasound signal from pressure data from a single sensor. The spatial average of the pressure field from the multiple array elements reduced the wind noise and revealed the infrasound signal for winds below about 2 m/s. During the test in high wind conditions, the resultant wind noise at the sensors masked the arrival of the blast-induced infrasonic signal on both the single sensor and spatial average of the array. For all four tests, calculating the magnitude of the signal-to-noise ratio for each time step identified the arrival of the infrasonic signal during both low and high noise conditions. These results compared favorably to F-statistic calculations for successfully identifying the blast-induced infrasound events from all four tests.

[1]C. R. Wilson and R. B. Forbes, "Infrasonic waves from Alaskan volcanic eruptions," J. Geophys. Res. **74**, 4511–4522 (1969).
[2]M. Garcés, M. Iguchi, K. Ishihara, M. Morrissey, Y. Sudo, and T. Tsutsui, "Infrasonic precursors to a vulcanian eruption at Sakurajima volcano, Japan," Geophys. Res. Lett. **26**, 2537–2540 (1999).
[3]J. B. Johnson, R. C. Aster, and P. R. Kyle, "Volcanic eruptions observed with infrasound," Geophys. Res. Lett. **31**, L14604 (2004).
[4]W. C. Meecham, "Effect of atmospheric wind structure on shorter-period, nuclear-generated infrasound," J. Geophys. Res. **73**, 377–384 (1968).
[5]D. R. Christie, B. L. N. Kennett, and C. Tarlowski, "Detection of atmospheric explosions at IMS monitoring stations using infrasound techniques," in 28th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies (2006).
[6]A. Le Pichon, P. Herry, P. Mialle, J. Vergoz, N. Brachet, M. Garcés, D. Drob, and L. Ceranna, "Infrasound associated with 2004–2005 large Sumatra earthquakes and tsunami," Geophys. Res. Lett. **32**, L19802 (2005).
[7]A. Le Pichon, J. Guilbert, M. Vallée, J. X. Dessa, and M. Ulziibat, "Infrasonic imaging of the Kunlun Mountains for the great 2001 China earthquake," Geophys. Res. Lett. **30**, 1814 (2003).
[8]M. Garcés, P. Caron, C. Hetzer, A. Le Pichon, H. Bass, D. Drob, and J. Bhattacharyya, "Deep infrasound radiated by the Sumatra earthquake and tsunami," EOS Trans. Am. Geophys. Union **86**, 317 (2005).
[9]J. Aucan, D. Fee, and M. Garcés, "Infrasonic estimation of surf period," Geophys. Res. Lett. **33**, L05612 (2006).
[10]M. Garcés, J. Aucan, D. Fee, P. Caron, M. Merrifield, R. Gibson, J. Bhattacharyya, and S. Shah, "Infrasound from large surf," Geophys. Res. Lett. **33**, L05611 (2006).
[11]A. Le Pichon, V. Maurer, D. Raymond, and O. Hyvernaud, "Infrasound from ocean waves observed in Tahiti," Geophys. Res. Lett. **31**, L19103 (2004).
[12]S. J. Arrowsmith and M. A. H. Hedlin, "Observations of infrasound from surf in southern California," Geophys. Res. Lett. **32**, L09810 (2005).
[13]C. R. Wilson, J. V. Olson, and H. C. Stenbaek-Nielsen, "High trace-velocity infrasound from pulsating auroras at Fairbanks, Alaska," Geophys. Res. Lett. **32**, L14810 (2005).
[14]T. Farges, E. Blanc, A. Le Pichon, T. Neubert, and T. H. Allin, "Identification of infrasound produced by sprites during the Sprite2003 campaign," Geophys. Res. Lett. **32**, L01813 (2005).
[15]P. G. Brown, R. W. Whitaker, D. O. ReVelle, and E. Tagliaferri, "Multi-

station infrasonic observations of two large bolides: Signal interpretation and implications for monitoring of atmospheric explosions," Geophys. Res. Lett. **29**, 1636 (2002).
[16]L. G. Evers and H. W. Haak, "Tracing a meteoric trajectory with infrasound," Geophys. Res. Lett. **30**, 2246 (2003).
[17]M. Garcés, M. Bass, D. Drob, C. Hetzer, H. Hedlin, A. Le Pichon, K. Lindquist, R. North, and J. Olson, "Forensic studies of infrasound from massive hypersonic sources," EOS Trans. Am. Geophys. Union **85**, 433 (2004).
[18]S. Morgan and R. Raspet, "Investigation of the mechanisms of low-frequency wind noise generation outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).
[19]R. Raspet, J. Webster, and K. Dillion, "Framework for wind noise studies," J. Acoust. Soc. Am. **119**, 834–843 (2006).
[20]F. D. Shields, "Low-frequency wind noise correlation in microphone arrays," J. Acoust. Soc. Am. **117**, 3489–3496 (2005).
[21]M. A. H. Hedlin, B. Alcoverro, and G. D'Spain, "Evaluation of rosette infrasonic noise reducing filters," J. Acoust. Soc. Am. **114**, 1807–1820 (2003).
[22]T. Marston and T. Gabrielson, "Modeling real porous tube infrasonic arrays," J. Acoust. Soc. Am. **123**, 3839 (2008).
[23]K. T. Walker, M. A. Zumberge, M. A. H. Hedlin, and P. M. Shearer, "Methods for determining infrasound phase velocity direction with an array of line sensors," J. Acoust. Soc. Am. **124**, 2090–2099 (2008).
[24]J. V. Olson, "Infrasound signal detection using the Fisher F-statistic," Inframatics, No. 6, pp. 1–7 (June 2004).
[25]*CRC Standard Mathematical Tables and Formulae*, 30th ed., edited by D. Zwillinger (CRC, Boston, MA, 1996).
[26]*Theory and Problems of Statistics*, 3rd ed., edited by M. R. Speigel and L. J. Stephens (McGraw-Hill, New York, 1999).
[27]B. S. Melton and L. F. Bailey, "Multiple signal correlators," Geophysics **22**, 565–588 (1957).
[28]S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing* (California Technical, San Diego, CA, 1997).
[29]E. T. Herrin, P. W. Golden, H. E. Bass, D. E. Norris, B. Andre, C. de Groot-Hedlin, R. L. Woodward, K. T. Walker, D. P. Drob, C. A. L. Szuberla, M. A. H. Hedlin, R. W. Whittaker, M. A. Garcés, and F. D. Shields, "High-altitude infrasound calibration experiments," Acoust. Today **4**, 9–21 (2008).

A multi-mode screech frequency prediction formula for circular supersonic jets

J. H. Gao and X. D. Li[a)]
School of Jet Propulsion, Beihang University, Beijing 100191, People's Republic of China

(Received 13 July 2009; revised 16 December 2009; accepted 17 December 2009)

A time evolution analysis is presented for the interaction between the instability waves, shock cells, and screech tones based on the authors' previous numerical simulation database. An attachment and reinforcement process of the upstream propagating screech waves with the downstream hydrodynamic waves is identified and recognized as part of the screech loop. The first five shock cells are recognized as the effective sound source region. Through an analysis of the phase variation in the dominant pressure fluctuations for several typical Mach number screeching jets, it is found that the total number of the instability waves and the upstream feedback sound waves in the effective source region can be identified as 5 for the A1, B, and D modes and 6 for A2 and C modes, respectively. A screech tone frequency prediction formula is thus proposed based on this relation. The predicted screech wavelengths or Strouhal numbers of cold and hot jets all agree well with the experimental data by other researchers, except for a small discrepancy for the B mode. It is also noticed that the measured two A0 modes by Ponton *et al.* [NASA Technical Memorandum No. 113137, Langley Research Center (1997)] can be classified to A1 and A2 modes, respectively, according to the proposed formula.
 [DOI: 10.1121/1.3291001]

PACS number(s): 43.28.Ra [AH] Pages: 1251–1257

I. INTRODUCTION

Jet screech has long been a fascinating subject of aeroacoustic research since Powell's[1] first observation in 1951. A large number of investigations on jet screech have been conducted from various aspects, and great advances have been achieved for the understanding of the sound generation mechanism which can be found in the reviews by Raman.[2,3] Among various investigations, the accurate prediction of the frequency and amplitude of jet screech tones has been one of the most important topics. The pioneering experimental observation of Powell elucidated jet screech as a nonlinear feedback mechanism. This provided a basis for his phased monopole model,[1] which explained the highly directional nature of screech. Tam *et al.*[4] also proposed a formula for screech frequency prediction. This formula can predict the screech frequencies of circular jet well at high Mach numbers, but there is a deviation for those modes at low supersonic Mach number. Panda[5] investigated the screech generation processes and provided detailed information on the time evolution of near-field pressure fluctuations. Based on the observation of a standing wave pattern in his experiment, Panda[5] established the relationship between the standing wave, the hydrodynamic wave, and the screech tone with the assumption of free acoustic reflections and proposed a screech tone frequency prediction formula. The standing wave pattern is a resultant of the upstream propagating screech waves with the downstream propagating hydrodynamic waves, and there is no explicit way to calculate the wavelength of standing wave by any theoretical method unless the screech sound wavelength is known. Thus it is not straightforward to apply Panda's formula for screech frequency prediction.

With the rapid development of the computational aeroacoustics (CAA), many researchers used CAA method to study the screech tones.[6–11] Remarkable progress has been made by these researchers in prediction and understanding the screech tones. Recently, Li and Gao[9,10] simulated the axisymmetric and three dimensional (3D) round supersonic jet screech tones by a CAA method. Their calculated screech tone wavelengths and amplitudes over a wide range of Mach numbers agree well with the experimental data of Ponton *et al.*[12] The advantage of numerical simulation over experiment is that it can provide very detailed flow and acoustic field. As such, it is possible to provide very insightful information on screech generation mechanism.

The objective of this paper is to propose a more accurate frequency prediction method based on a detailed analysis of the large database provided in our previous numerical simulations.[9,10]

In Sec. II, detailed analysis of the screech generation and propagation process are presented and the generation mechanism is discussed. In Sec. III, a frequency prediction method based on the screech generation mechanism analysis is proposed and the predicted results are compared with the experimental data by other researchers. Section IV gives the conclusions of this work.

II. FLOW AND SOUND NEAR FIELD ANALYSIS

A. Time evolution of the screech generation

The axisymmetric and full 3D supersonic round jet screech tones were simulated by the present authors in their

a)Author to whom correspondence should be addressed. Electronic mail: lixd@buaa.edu.cn

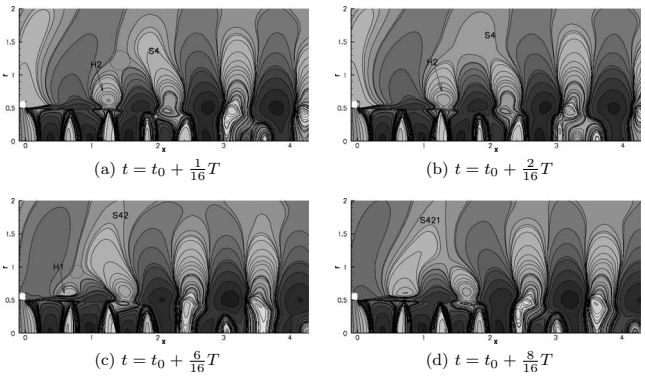


FIG. 1. Instantaneous pressure field of M_j=1.10 screeching jet at different times showing the interaction of screech with hydrodynamics waves.

previous work[9,10] by a CAA method. In those simulations, a fourth order dispersion-relation-preserving (DRP) scheme[13] was used for space discretization. A 5/6-stage low-dissipation and low-dispersion Runge–Kutta scheme[14] in the 2*N* storage form was employed for time integration. The detailed flow and near sound field of jets in a large Mach number range was obtained. Based on the database an analysis of the screech generation process is presented below.

A time evolution of the screech generation, propagation, and interaction with the hydrodynamic waves in the near sound field for the M_j=1.10 screeching jet is shown in Fig. 1. As shown in Fig. 1(a), the crest of a screech wave, labeled as S4, generated between the third and fourth shock cells, begins to attach to the crest of the downstream convecting hydrodynamic wave (labeled as H2) near the second shock cell. The S4 wave is then successively reinforced by H2 until it is detached from H2, as shown in Fig. 1(b). The reinforced screech wave, labeled as S42, and the hydrodynamic waves H2 continue to move upstream and downstream, respectively. The wave S42 propagating upstream begins to attach to and be reinforced by the hydrodynamic wave H1, which can be observed clearly in Fig. 1(c). In Fig. 1(d), when S42 attached completely to H1, the crest of the reinforced screech wave, labeled as S421, propagates upstream, reaches the nozzle exit, and excites a new instability wave in the near nozzle region due to receptivity.

A similar process can be found for the troughs of the screech waves and the troughs of the hydrodynamic waves, which are not presented here. It can also be observed in these figures that the strongest screech sound waves are generated between the second and fourth shock cells. Only weak sound waves are emitted in the regions in front of the first shock and beyond the fourth shock. The above analysis shows that the first five shock-cell spacing can be recognized as the effective source region. Similar but more complicated process can also be observed for the flapping and helical screeching jets, which cannot be presented here for brevity. The phenomenon described above is very similar to that observed in a lot of experiments by Davies and Oldfield,[15] Westley and Woolley,[16] Chan,[17] Panda,[5] and so on.

B. Phase variation

Panda[5] and Umeda and Ishii[18] studied the phase variation of the screech sound waves in their experiments, but

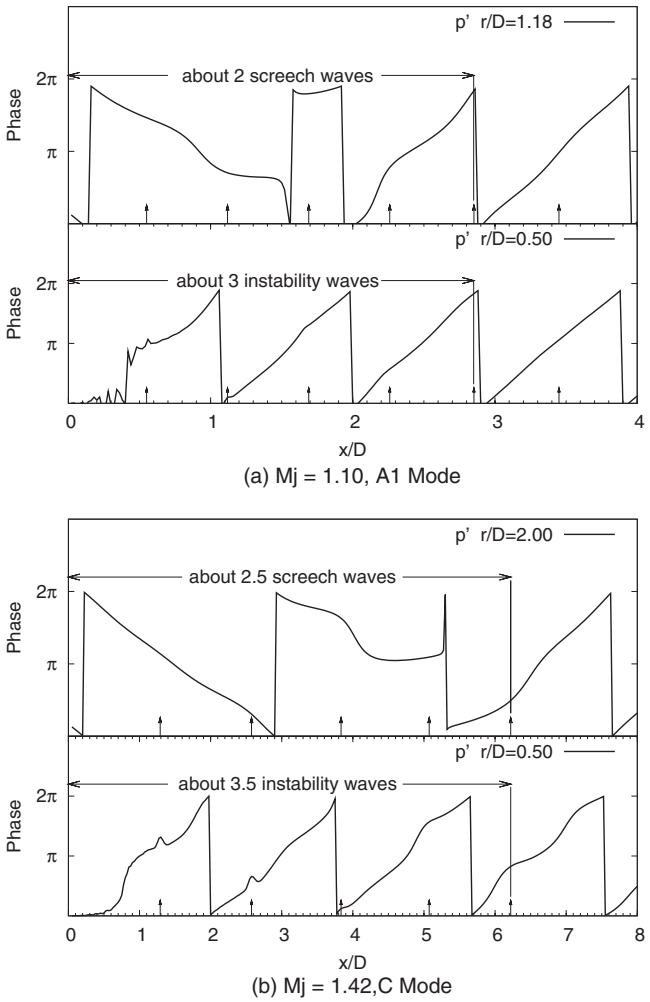


FIG. 2. Phase variation of the dominant mode pressure fluctuations in the shear layer for the M_j=1.10 and 1.42 screeching jets. The vertical arrows at the bottom of the figures represent shock locations.

they did not associate it with the screech frequency prediction. In this paper, a phase variation analysis similar to Panda[5] is applied to our numerical database. To obtain the phase of the pressure fluctuations in the shear layer and the near field outside the jet, cross-correlation functions between the signal at point near the nozzle lip (reference point) and the signals at other points are first calculated. Through a Fourier transformation of the cross-correlation functions, the relative phase (ϕ) of the dominant mode ($f=f_s$) in the shear layer and outside the jet is determined. The phase variations of pressure fluctuations in dominant mode along the x direction at fixed radial positions for several typical screeching jets are analyzed. Only the results of jets with Mach numbers 1.10 and 1.42 are plotted in Figs. 2(a) and 2(c), respectively, for limited space. In each figure the black vertical arrows at the bottom represent the shock locations. For the axisymmetric M_j=1.10 screeching jet, the number of waves along r/D=0.5 line in the first five shock-cell spacing is roughly 3, as shown in Fig. 2(a). Since the relative phase along r/D =0.5 increases with the axial positions, and the hydrodynamic waves dominate others along r/D=0.5, it can be considered that the phase along r/D=0.5 represents the phase of the hydrodynamic waves. In the upper part of Fig. 2(a), the

relative phase along $r/D=1.18$ decreases first and then increases. The decreasing part represents the phase of the upstream propagating sound waves, and the increasing part represents the phase of the hydrodynamic waves. With the help of the numerical predicted screech wavelength by CAA, the number of sound waves in the first five shock-cell spacing can be calculated and it is roughly 2. The total number of the instability waves and screech sound waves in the effective source region for the axisymmetric $M_j=1.10$ jet is 5. In this way, the total number of waves can be determined as 5 for the jets in flapping mode and 6 for jets in helical mode, which is shown in Fig. 2(c) for the $M_j=1.42$ jet.

It is known that the travel time associated with the convection of the downstream traveling hydrodynamic disturbances plus the time taken by the upstream traveling acoustic disturbance sum up to an integral number of periods of the screech cycle.[19] If the time delay in the receptivity and sound generation is not accounted for, the following equation can be established:

$$\frac{nL}{u_c}+\frac{nL}{c}=mT, \tag{1}$$

where T is the period of the screech cycle, m is an integral number and represents the number of screech cycles, u_c is the instability convective velocity, c is the ambient sound speed, L is the averaged shock-cell spacing, nL is the length of the effective source region, and $n=5$ according to observation.

Equation (1) can be rewritten as

$$\frac{nL}{u_cT}+\frac{nL}{cT}=m,$$

$$\frac{nL}{\lambda_h}+\frac{nL}{\lambda_s}=m, \tag{2}$$

where λ_h is the instability wavelength and λ_s is the screech sound wavelength. This equation means that the total number of the instability wave and the screech sound wave in the effective source region is equal to an integer. From the above wave number analysis it is known that for the axisymmetric $M_j=1.10$ screeching jet, $m=5$, for the $M_j=1.30$ flapping screeching jet, $m=5$, but for the $M_j=1.42$ helical screeching jet, $m=6$. Associated the value variation of m with the mode staggering from A1 to D mode, it is straightforward to assume that for the A1, B, and D modes, the value of m is 5 and for the A0, A2, and C modes, it is 6. This can be summarized as

$$\begin{cases} m=5 & \text{for A1, B, D mode} \\ m=6 & \text{for A0, A2, C mode.} \end{cases} \tag{3}$$

This assumption will be verified by the prediction results in the latter part of this paper. This equation relates the screech mode to the total number of instability waves and the screech feedback waves in a screech cycle. If the total number of waves in a whole cycle changes, the screech mode would change. This may be the reason of the mode switching for the round supersonic jet screech tones.

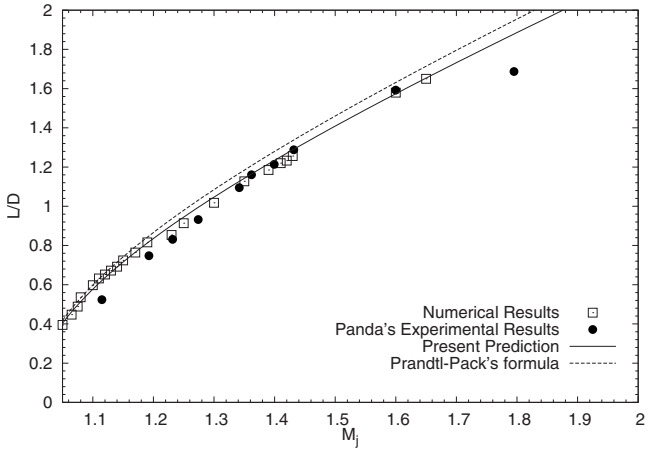


FIG. 3. Predicted shock-cell spacing with Prandtl–Pack formula compared with the experimental data of Panda *et al.* (Ref. 24) and the CAA results (Refs. 9 and 10).

III. FREQUENCY PREDICTION

Based on Eq. (2) and the relationship of $\lambda_h=\lambda_s M_c$, a formula for screech tone wavelength prediction can be obtained as

$$\lambda_s=\frac{nL}{m}\frac{1+M_c}{M_c}, \tag{4}$$

where M_c is the convective Mach number of the instability waves. Coupled with the assumption [Eq. (3)] of the total wave number in the effective source region for each mode, this new formula can be used to predict the screech wavelength if the averaged shock-cell spacing and the convective Mach number are known.

A. Shock spacing

Shock spacing is important in predicting jet screech tones. Several researchers studied the shock-cell spacing.[20–23] The Prandtl–Pack[20] formula provides a simple relationship between the shock spacing and the jet Mach number M_j based on a non-dissipative analysis:

$$L/D=1.306\sqrt{M_j^2-1}, \tag{5}$$

where D is the jet diameter. According to the study of Harper-Bourne and Fisher,[23] the Prandtl–Pack formula can predict well the first shock-cell length from the nozzle. But it cannot account for the shock-cell length reduction along the jet due to viscous dissipation. Therefore, the average shock spacing predicted by Prandtl–Pack formula would be greater than the experimental data[24] and the CAA results.[9,10] This is shown clearly in Fig. 3. Based on our previous numerical simulation results[9,10] and the experimental data of Panda *et al.*,[24] the coefficient in Prandtl–Pack formula is modified and calibrated. The formula with the modified coefficient is

$$L/D=1.26\sqrt{M_j^2-1}. \tag{6}$$

As shown in Fig. 3, the predicted shock spacing using the formula with modified coefficient agrees much better with the experimental data of Panda *et al.*[24] than that with the original coefficient.

TABLE I. Convective Mach number (or velocity) coefficient provided by other researchers [where data are u_c/u_j value].

	A mode	B mode	C mode	D mode
Powell *et al.*[a]	0.64	0.68	0.8	0.75
Panda *et al.*[b]	0.67	0.58	0.66	0.69
Tam *et al.*[c]			All 0.7	

[a]Reference 26.
[b]Reference 24.
[c]Reference 4.

B. Convective Mach number

Based on the assumption that the ratio of the convective velocity to the axial velocity is a fixed value, Tam *et al.*[4] proposed a formula to calculate the convective Mach number:

$$M_c = u_c/c = Cu_j/c = CM_j\left(1+\frac{\gamma-1}{2}M_j^2\right)^{-1/2}\sqrt{\frac{T_r}{T_a}}, \quad (7)$$

where C is a constant coefficient, M_j is the jet fully expanded Mach number, T_r/T_a is the ratio of the jet reservoir temperature T_r to the ambient air temperature T_a, and γ is the specific heat ratio. Several researchers have provided the convective Mach number (or velocity) data, as shown in Table I. In this table, the convective Mach number coefficient (C) data by different researchers are not consistent. This indicates that it is difficult to predict the convective Mach number accurately, especially when it varies with the screech modes. Panda[5] suggested a formula to calculate the convective Mach number with the phase derivative of instability waves along the shear layer:

$$M_c = \frac{u_c}{c} = \frac{2\pi f_s}{cd\phi/dx}. \quad (8)$$

Because the phase derivative in the axial direction is very large at some points (e.g., near the shocks), the calculated convective Mach number varies wildly. It is hard to calculate an averaged convective Mach number accurately by this method. Since an averaged convective Mach number in the effective source region is adequate for the screech wavelength prediction, the averaged form of Panda's method is applied here:

$$\bar{M}_c = \frac{\Delta x}{c\Delta t} = \frac{\Delta x}{c\Delta\phi/(2\pi f_s)} = 2\pi\frac{\Delta x}{\lambda_s\Delta\phi}, \quad (9)$$

where f_s is the screech tone frequency and $\Delta\phi$ is the phase variation in the distance of Δx. The relative phase variations of the dominated pressure fluctuations in the shear layer are calculated for M_j=1.10, 1.30, and 1.42 screeching jets, and only the results of M_j=1.30 screeching jet are plotted in Fig. 4. In this figure, a distance of two hydrodynamic wavelengths for Δx is used for the convenience of measuring the corresponding phase variation. Only the latter part of the phase variations in the effective source region is considered when calculating the averaged convective Mach number. In the region near the nozzle lip wall, the acoustic waves are dominant and a standing wave is formed by the upstream propagating sound wave and the reflected sound wave on the wall; thus it is difficult to determine the phase variation of the instability waves accurately. According to the measured distances and the corresponding relative phase variations, the calculated averaged convective Mach numbers are 0.572, 0.69, and 0.737 for Mach numbers 1.10, 1.30, and 1.42 screeching jets, respectively. According to Eq. (7), the coefficient C in the convective Mach number prediction formula can be determined as 0.58, 0.614, and 0.615 for the three cases, respectively. Therefore, the following coefficients will be used for the following screech prediction:

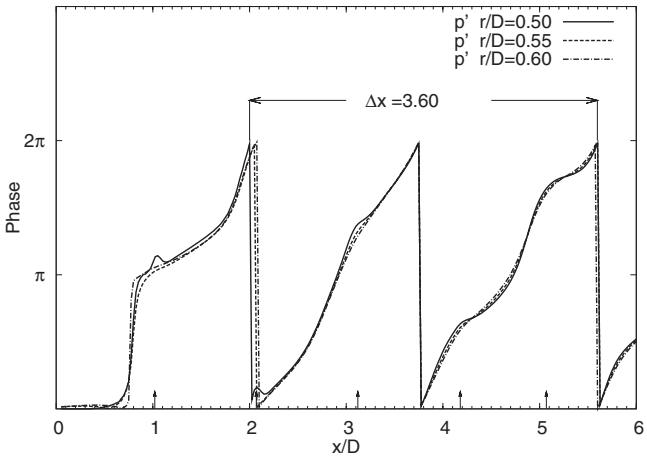


FIG. 4. Phase variation of the dominant mode pressure fluctuations in the shear layer for the M_j=1.30 screeching jet. The vertical arrows at the bottom of the figures represent shock locations.

$$C = \begin{cases} 0.580 & \text{for axisymmetric mode} \\ 0.615 & \text{for flapping and helical mode.} \end{cases} \quad (10)$$

It should be noted that a value of 0.615 is also used for jets with Mach number greater than 1.2 by Massey[25] to predict the screech frequency using Tam *et al.*'s formula.

C. Screech wavelength prediction

Based on Eqs. (4), (6), and (7) and the coefficients (3) and (10), the supersonic circular jet screech tone wavelength can be predicted.

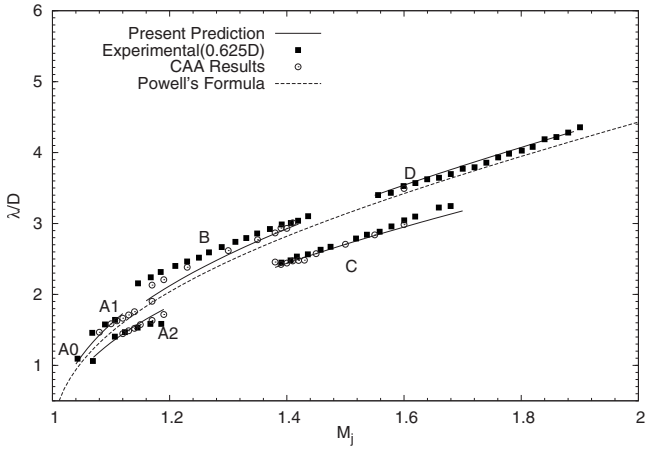


FIG. 5. Predicted screech tone wavelength with the present formula compared with the experimental data of Ponton *et al.* (Ref. 12) and that with Powell's formula.

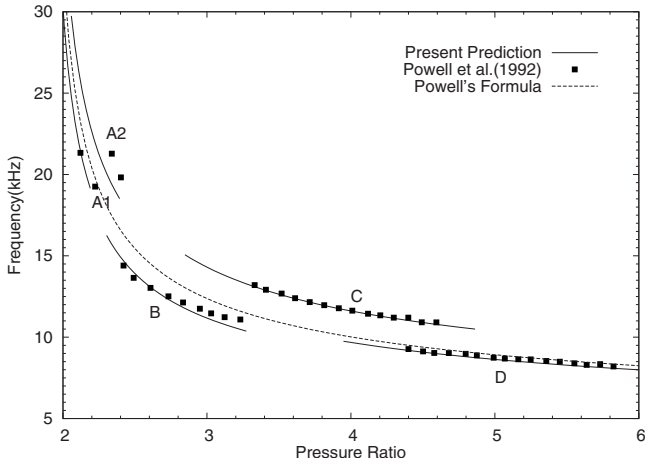


FIG. 6. Predicted screech frequencies versus measurements of Powell *et al.* (Refs. 18 and 26).

The predicted wavelengths for the supersonic cold jets in a large Mach number range are shown in Fig. 5. The predicted results are compared with the experimental data of Ponton *et al.*[12] and that with Powell's[19] formula. It is shown that the present predicted results agree well with the experimental data for A1, A2, C, and D modes, except for a small deviation for B mode. However, the predicted results with Powell's[19] formula only agree well with the experimental data for the A1 and D modes. It is interesting to find that the measured two A0 modes in the experiment of Ponton *et al.*[12] can be classified into A1 and A2 modes, respectively, according to the predicted results. In Fig. 6, the predicted dominant screech frequencies are presented and compared with the experimental data of Powell *et al.*[18,26] and that with Powell's[19] formula. It is shown that the predicted frequencies with the present formula agree well with the experimental data except for a small discrepancy for the A2 mode. However, the predicted results with Powell's formula only agree well with the experimental data for the A1 and D modes. In Fig. 7, the present predicted results are also compared with the experimental data of Chen *et al.*[27] and Yu[28] (quoted in the paper of Chen *et al.*[27]) and that with Tam *et al.*'s formula. The results with the present formula agree well with the experimental

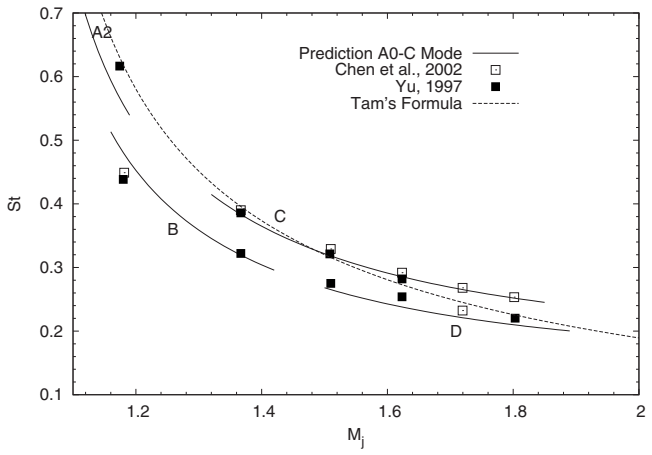


FIG. 7. Predicted screech Strouhal numbers versus measurements of Chen *et al.* (Ref. 27) and Yu (Ref. 28).

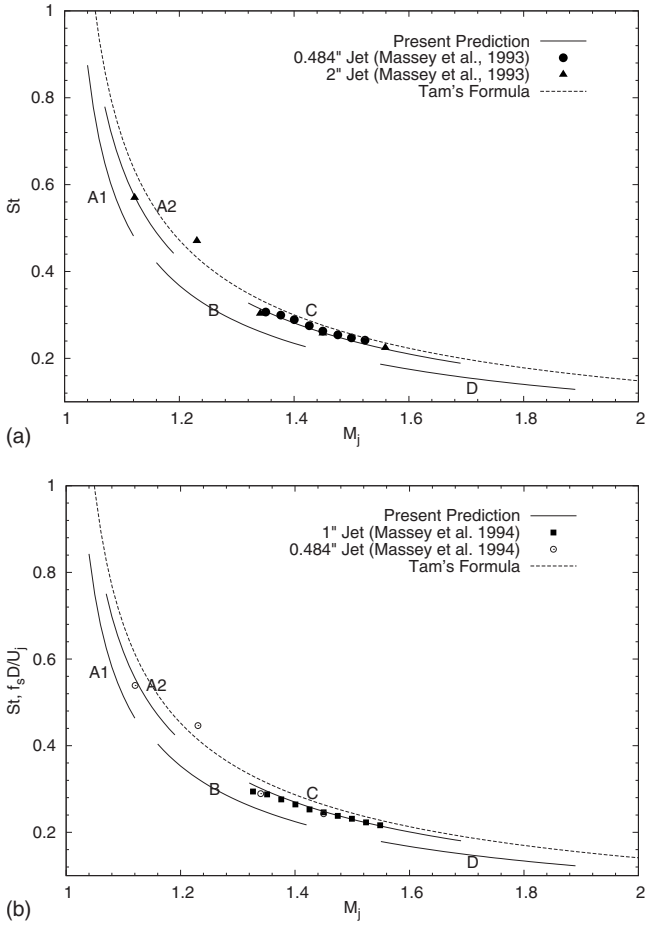


FIG. 8. Predicted screech Strouhal number versus measurements (Ref. 29) for jet with temperature ratios 2.37 and 2.78.

data except for a deviation for the point in A2 mode. Also, Tam *et al.*'s formula can only predict well for a limited Mach number range. In Fig. 8 the predicted Strouhal numbers for the hot jets are compared with the experimental data of Massey *et al.*[29] and that with Tam *et al.*'s[4] formula. The Strouhal number is calculated by the following equation:

$$\mathrm{St} = \frac{f_s D}{u_j} = \frac{m}{nL}\frac{CD}{1+M_c}. \tag{11}$$

It is shown that for the jets with temperature ratios of 2.37 and 2.78, the predicted Strouhal numbers agree better with the experimental data than the prediction results with Tam *et al.*'s formula. In Fig. 9, the predicted Strouhal numbers for the jet with different temperature ratios and Mach numbers are compared with the experimental data by Krothapalli *et al.*[30] and the predicted results with Tam *et al.*'s[4] formula. Here the Strouhal number is defined with the fully expanded jet diameter D_j instead of the jet diameter D.

$$\frac{D_j}{D} = \left\{\frac{1+[(\gamma-1)/2]M_j^2}{1+[(\gamma-1)/2]M_d^2}\right\}^{(\gamma+1)/4(\gamma-1)}\left(\frac{M_d}{M_j}\right)^{1/2}, \tag{12}$$

where M_d is the nozzle design Mach number. For the convergent circular nozzle, M_d=1. It should also be noted that for the results predicted with Tam *et al.*'s formula, the value of 0.615 for the coefficient C is used in calculation of the convective Mach number M_c. Since Tam *et al.*'s[4] formula

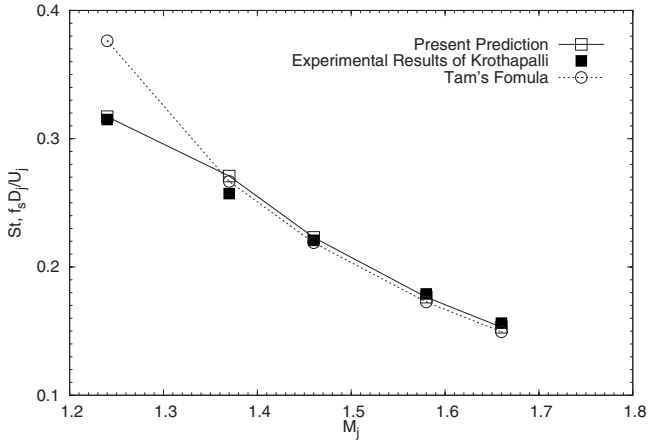


FIG. 9. Predicted screech Strouhal number compared with measurements (Ref. 30) and Tam *et al.*'s (Ref. 4) formula for jets with temperature ratios 2.5–3.5 and Mach numbers 1.24–1.66

considers all screech tones as a single mode, there is a large deviation for the first point. Actually, the mode of the first point should be B and the others should be C mode. The good agreement between the predicted results and the experimental data demonstrates that the assumption [Eq. (3)] of the total wave number in the effective source region for each mode is viable.

IV. CONCLUSIONS

Based on our previous numerical simulation database, the time evolution of the interaction between the instability waves, shock cells, and screech tones is analyzed and presented. The attachment and reinforcement process of the upstream propagating screech waves with the hydrodynamic waves is identified and recognized as part of the screech cycle. The first five shock cells are recognized as the effective source region, and the strongest screech waves are generated between the second and fourth shock cells. Through an analysis of the phase variation of the dominant pressure fluctuations for several typical Mach number screeching jets, it is found that the total number of instability waves and the feedback screech sound waves in the effective source region can be identified as 5 for the A1, B, and D modes and 6 for A2 and C modes, respectively. A new screech tone wavelength prediction formula is proposed based on this relation. A thorough comparison is done for the prediction results by this formula with the experimental data by Ponton *et al.*,[12] Powell *et al.*,[18,26] Chen *et al.*,[27] Yu,[28] Massey *et al.*,[29] and Krothapalli *et al.*,[30] respectively. The predicted results agree well with all the presented experimental data except for a small discrepancy for minority points. It is also found that the measured two A0 modes in the experiment by Ponton *et al.*[12] can be classified to A1 and A2 modes, respectively, according to the proposed prediction formula. This implies that there is no A0 mode for the round supersonic jet screech tones in reality.

ACKNOWLEDGMENTS

This work was supported by grants from NSFC Grant No. 10802004, NCET Grant No. 07-0038, and 973 Program Grant No. 2007CB714604. The authors would also like to acknowledge the partial support from the 111 projects B07009 and B08009 of China and the computing time from the Aeroengine Numerical Simulation Research Center (ASRC) in BUAA.

[1]A. Powell, "On the noise emanating from a two-dimensional jet above the critical pressure," Aeronaut. Q. **4**, 103–122 (1953).
[2]G. Raman, "Advances in understanding supersonic jet screech: Review and perspective," Prog. Aerosp. Sci. **34**, 45–106 (1998).
[3]G. Raman, "Supersonic jet screech: Half-century from Powell to the present," J. Sound Vib. **225**, 543–571 (1999).
[4]C. K. W. Tam, J. M. Seiner, and J. C. Yu, "Proposed relationship between broadband shock associated noise and screech tones," J. Sound Vib. **110**, 309–321 (1986).
[5]J. Panda, "An experimental investigation of screech noise generation," J. Fluid Mech. **378**, 71–96 (1999).
[6]H. Shen and C. K. W. Tam, "Numerical simulation of the generation of axisymmetric mode jet screech tones," AIAA J. **36**, 1801–1807 (1998).
[7]H. Shen and C. K. W. Tam, "Three-dimensional numerical simulation of the jet screech phenomenon," AIAA J. **40**, 33–41 (2002).
[8]T. A. Manning and S. K. Lele, "Numerical simulation of shock-vortex interactions in supersonic jet screech," AIAA Paper No. 1998-0282 (1998).
[9]X. D. Li and J. H. Gao, "Numerical simulation of the generation mechanism of axisymmetric supersonic jet screech tones," Phys. Fluids **17**, 085105 (2005).
[10]X. D. Li and J. H. Gao, "Numerical simulation of the three dimensional screech phenomenon from a circular jet," Phys. Fluids **20**, 035101 (2008).
[11]J. Berland, C. Bogey, and C. Bailly, "Numerical study of screech generation in a planar supersonic jet," Phys. Fluids **19**, 075105 (2007).
[12]M. K. Ponton, J. M. Seiner, and M. C. Brown, "Near field pressure fluctuations in the exit plane of a choked axisymmetric nozzle," NASA Technical Memorandum No. 113137, Langley Research Center (1997).
[13]C. K. W. Tam and J. C. Webb, "Dispersion-relation-preserving finite difference schemes for computational acoustics," J. Comput. Phys. **107**, 262–281 (1993).
[14]D. Stanescu and W. G. Habashi, "2N-storage low-dissipation and low-dispersion Runge-Kutta schemes for computational aeroacoustics," J. Comput. Phys. **143**, 674–681 (1998).
[15]M. G. Davies and D. E. S. Oldfield, "Tones from a choked axisymmetric jet," Acustica **12**, 257–277 (1962).
[16]R. Westley and J. H. Woolley, "The near field sound pressure of a chocked jet during a screech cycle," in The AGARD Conference on Aircraft Engine Noise and Sonic Boom (1969), AGARD CP-42, pp. 23-1– 23-13.
[17]Y. Y. Chan, "A simple model of shock-cell noise generation and its reduction," National Research Council/National Aeronautical Establishment (Canada) Aeronautics Report No. LR 564, 1972.
[18]Y. Umeda and R. Ishii, "On the sound sources of screech tones radiated from choked circular jets," J. Acoust. Soc. Am. **110**, 1845–1858 (2001).
[19]A. Powell, "On the mechanism of choked jet noise," Proc. Phys. Soc. London **66**, 1039–1056 (1953).
[20]D. C. Pack, "A note on Prandtl's formula for the wavelength of a supersonic gas jet," Q. J. Mech. Appl. Math. **3**, 173–181 (1950).
[21]A. Powell, "The focal phase shift of waves of a circular membrane, applied to underexpanded supersonic jet structure," J. Acoust. Soc. Am. **97**, 927–932 (1995).
[22]C. K. W. Tam, J. A. Jackson, and J. M. Seiner, "A multiple scales model of the shock-cell structure of imperfectly expanded supersonic jets," J. Fluid Mech. **153**, 123–149 (1985).
[23]M. Harper-Bourne and M. J. Fisher, "The noise from shock waves in supersonic jets," in The AGARD Conference on Noise Mechanisms (1973), AGARD-CP-131, pp. 1–13.
[24]J. Panda, G. Raman, and K. Zaman, "Underexpanded screeching jets from circular, rectangular and elliptic nozzles," AIAA Paper No. 1997-1623 (1997).
[25]K. C. Massey, "Flow/acoustic coupling in heated and unheated free and ducted jets," Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA (1997).
[26]A. Powell, Y. Umeda, and R. Ishii, "Observation of oscillation modes of choked circular jets," J. Acoust. Soc. Am. **92**, 2823–2836 (1992).

[27]R. H. Chen, S. B. Saadani, and L. P. Chew, "Effects of nozzle size on screech noise elimination from swirling underexpanded jets," J. Sound Vib. **252**, 178–186 (2002).
[28]Y.-K. Yu, "An experimental study of underexpanded supersonic swirling jets," Ph.D. thesis, University of Central Florida, Orlando, FL (1997).
[29]K. C. Massey, K. K. Ahuja, R. R. Jones, and C. K. W. Tam, "Screech tones of supersonic heated free jets," AIAA Paper No. 1994-0141 (1994).
[30]A. Krothapalli, P. T. Soderman, C. S. Allen, J. A. Hayes, and S. M. Jaeger, "Flight effects on the far-field noise of a heated supersonic jet," AIAA J. **35**, 952–957 (1997).

Application of the coherent-to-incoherent intensity ratio to estimation of ocean surface roughness from high-frequency, shallow-water propagation measurements

Philippe Roux
LGIT, CNRS UMR 5559, Université Joseph Fourier, 1381 rue de la Piscine, 38041 Grenoble, France

R. Lee Culver
Applied Research Laboratory and Graduate Program in Acoustics, Penn State University, State College, Pennsylvania 16804

Shane Walker
Marine Physical Laboratory, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92038

(Received 24 March 2009; revised 14 December 2009; accepted 27 December 2009)

For acoustic propagation through a shallow ocean channel or waveguide, the coherence between different transmissions is controlled primarily by the roughness of the ocean surface and to a lesser degree by fluctuations in the volume. In this study, the coherent-to-incoherent intensity ratio (CTIR) is defined as a way to quantify the coherence between multipath transmissions and ocean surface rms wave height and wind speed. A theory that connects the CTIR and the coherent surface reflection coefficient is developed using both Kirchhoff and small-slope approximations as rough surface scattering models. The CTIRs have been evaluated over a period of several days using broad-band experimental results from shallow-water deployment of source and receiver arrays that span most of the water column. Estimates of wind speed and rms wave height obtained using these CTIR calculations are compared with environmental measurements to demonstrate the validity of the theory. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3294493]

PACS number(s): 43.30.Hw, 43.30.Zk, 43.30.Gv [WLS] Pages: 1258–1266

I. INTRODUCTION

This study examines the relationships between the condition of the sea surface and the coherent average of multipath acoustic signals that have propagated through a shallow-water channel. Acoustic signal coherence is quantified via the *coherent-to-incoherent intensity ratio* (CTIR), a processing approach that has been used successfully recently to measure scattering by small objects in motion in reverberant cavities (de Rosny and Roux, 2001; de Rosny *et al.*, 2005). The present study derives a theory that relates the CTIR to the ocean surface conditions for propagation through a shallow ocean channel.

It is well known that reflection and scattering by a time-varying, rough surface, such as the surface of the ocean, decreases the coherent average of the acoustic field. When acoustic signals are forward scattered by a dynamic, stochastic, rough surface, the resulting field (after ensemble averaging) is composed of a coherent component that is due to a near-specular reflection process and an incoherent component that is due to surface scattering. Here *coherent* loosely refers to the acoustic wave component that is not changed from one realization to another, while *incoherent* refers to the randomly scattered wavefield for each realization. As surface roughness increases, the forward-scattered field becomes less coherent and more incoherent. In a shallow ocean water channel, numerous interactions of the acoustic signals with the sea surface reduce the coherent field proportionally to the number of surface reflections.

It was shown many years ago that acoustic scattering at a time-varying ocean surface causes ping-to-ping changes in total path length and thus the time of arrival of the received signal. For example, Lynch *et al.*, (1987) showed that surface-interacting acoustic paths in shallow water have "very prominent" travel-time fluctuations. Miller *et al.* (1989, 1993) used fluctuations in the travel times of surface-interacting paths to estimate the ocean surface-wave height spectrum. More recently, Lewis *et al.* (2005) used acoustic travel-time variations in a double-surface bounce path to obtain information about the sound speed in the ocean volume, to update an ocean circulation model; however, they found that the travel-time variations due to surface motion were dominant.

In the present study, temporal fluctuations of the received signal are quantified using the CTIR, where from Ishimaru, 1978 and Sheng, 1995, the coherent intensity is defined as the intensity of the ensemble-averaged received signal, and the incoherent intensity is defined as the ensemble average of the received intensity. Our experimental results show that the CTIR decreases exponentially with time to the 3/2 power after direct arrival and varies with the sea state in a predictable manner. The goal of this study is to determine whether the physics of the surface interaction together with CTIR processing can explain the observed time-dependent relationship.

From a theoretical and numerical point of view, a series of studies in the late 1980s provided useful descriptors of sea

 0001-4966/2010/127(3)/1258/9/$25.00

surface scattering using Kirchhoff and perturbation approximations (Ogilvy, 1987; Thorsos, 1988; Thorsos and Jackson, 1989). More recently, the small-slope approximation (SSA) reflection coefficient was shown to be the most accurate model for surface forward-scattering of acoustic waves (Broschat, 1993). Meanwhile, the Pierson–Moskowitz sea surface-wave height spectrum had been used successfully as a fully developed sea model (Pierson and Moskowitz, 1964, Thorsos, 1990). The SSA coherent surface reflection coefficient (CSRC) was then used to model the coherent pressure field scattered by the rough ocean surface (Williams *et al.*, 2004). Of note also, the spatial coherence of a sound field scattered once from the sea surface has been investigated in the time domain through the van Cittert–Zernike theorem (Dahl, 2004). Finally, the scintillation index was introduced from optics, as a way to characterize surface roughness in a bistatic configuration (Yang *et al.*, 1991).

Although the scattering of an acoustic wave on a rough sea surface has been extensively studied, most of the above-mentioned analyses are either (1) frequency-domain methods performed with continuous waves (at one or several discrete frequencies) in shallow waters or (2) time-domain approaches with only one bounce on the ocean surface. From this perspective, our definition of the CTIR serves as a complement to these methods, in that it combines an ensemble average of multireverberated broad-band acoustic arrivals in the time domain between a set of sources and a set of receivers that applies to the case of many boundary interactions. It appears intuitive that temporal shifts in the received signals will reduce the coherent intensity more than the incoherent intensity, and that travel-time fluctuations should increase with an increasing number of surface interactions.

This study is organized as follows: after this introduction, Sec. II deals with the measurement of the CTIR and with a description of the experimental configuration used in an at-sea measurement north of Elba Island, Italy, in April 2003. In Sec. III, the CTIR is used with both the Kirchhoff and the small-slope approximations to estimate rms wave height and wind speed. These estimates are then seen to compare favorably with wave-rider-buoy data and wind speed measurements in Sec. IV.

II. EXPERIMENTAL CONFIGURATION AND CTIR MEASUREMENT

We performed an at-sea experiment in April 2003 [Julian days (JDs) 101–106], north of Elba Island, Italy, using the equipment, location, and basic setup discussed by Roux *et al.* (2004). As shown in Fig. 1(a), 29 equally spaced source transducers spanned 78 m, and 32 equally spaced receiver hydrophones spanned 90 m, in water that was 120 m deep. The distance between the two arrays was 8.67 km. The bottom was composed of sand, and the sound speed profile varied by only 2 m/s over the water depth [Fig. 1(b)]. The data from each receiver hydrophone were transmitted to the support ship via radio frequency telemetry. Transmitted pulses were 100-ms, 1-kHz-bandwidth chirps, centered at F_c =3.5 kHz. The received signals were match filtered using

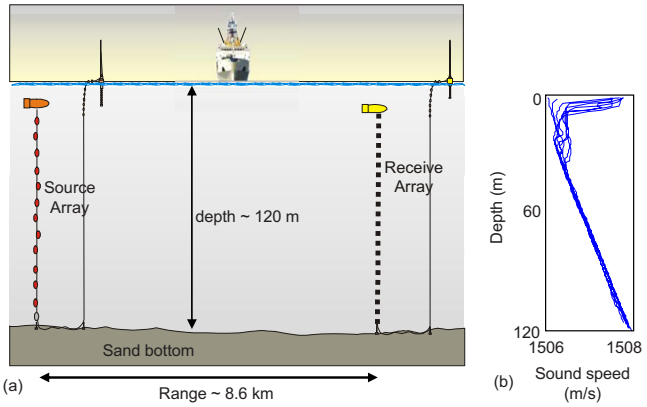


FIG. 1. (Color online) Experimental setup at sea. (a) The bottom-moored source array consisted of 29 transducers centered at 3.5 kHz that spanned 78 m of the water column. The bottom-moored receiver array consisted of 32 hydrophones with 90 m apertures. Both arrays were connected to surface buoys that contained batteries and radiofrequency telemetry hardware for data transfer and communication with the ship. The water depth was 118 m and the distance between the source and receiver arrays was 8.67 km. (b) Collection of sound speed profiles recorded in the area the same day that the acoustic experiment was performed. The average temperature profile is nearly uniform, leaving the sound speed profile with a constant pressure-related gradient.

the transmitted pulse to obtain an estimation of the channel impulse response with high signal-to-noise ratio and low transmission power.

The transmission setup used one acoustic projector transmitting at a time, while the received signal was recorded at all of the hydrophones (Fig. 2). The delay, T, between each transmission was chosen to be long enough for the signal to die out at all of the receiver hydrophones. The entire transmission sequence required 10 s, and it was repeated every 20 s over several long intervals (from 20 min to a few hours).

Figure 3 shows the matched filter output for the signals received on all of the hydrophones for a single transmission from the acoustic source at 58 m depth. As expected in a waveguide, the received signals are seen to be temporally dispersed, i.e., there is a direct path followed by arrivals that were reflected (or scattered) one or more times by the ocean surface and bottom. Time-delay beamforming was applied to the received pressure field in Fig. 3 to produce Fig. 4, which shows that the arrival angles increase with time after the first (direct path) arrival [Fig. 4(a)]. Distinct arrivals can be seen for up to 8° above and below the horizontal. There is good agreement with a ray-based acoustic prediction [Fig. 4(b)]. The ray paths plotted in Fig. 4(c) show that nearly all of the rays hit the ocean surface, a result that is due to the nearly uniform sound speed profile.

Figure 5 shows the signal arrival structure for a single source-receiver pair over a 7000 s (~2 h) period on JD 103. It is clear that the channel does not change significantly during this interval. However, it can be seen in Fig. 5 that late (multireverberated) signals lose coherence compared to (direct) early arrivals. Note also that the noise level after the direct path arrival is 5–10 dB higher than the noise level just prior to the direct path arrival. This is due to conversion of coherent acoustic energy to incoherent energy by sea surface scattering.

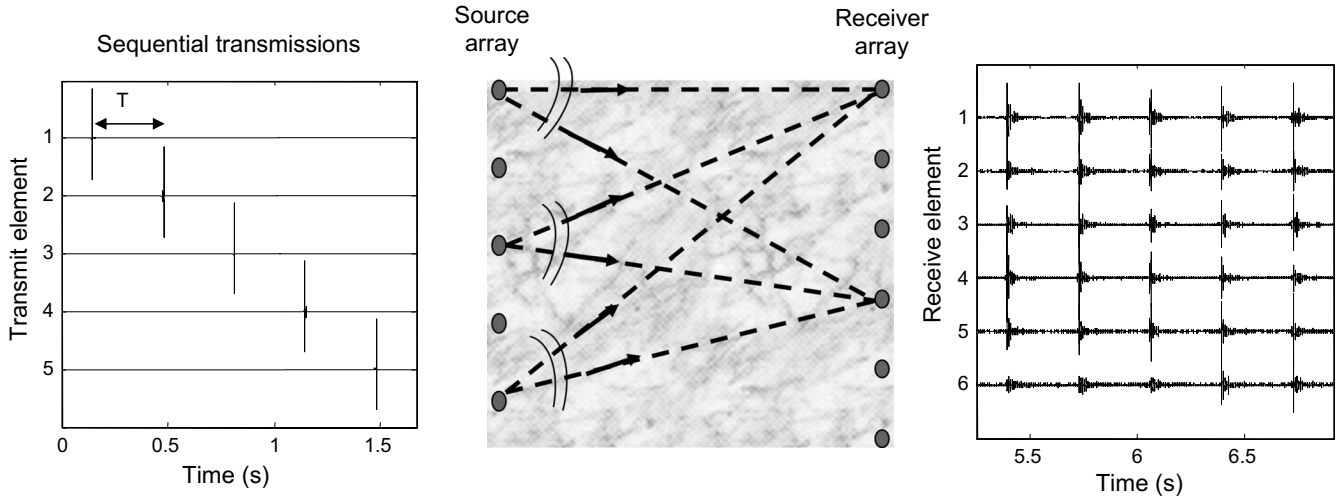


FIG. 2. Schematic of transfer matrix acquisition between the source and receiver arrays. The propagation medium between the two arrays can be any complex reverberating/scattering medium. (Left) The source array sequentially emits pulsed signals from each of its elements, separated by a time-window T that is larger than the expected acoustic dispersion in the propagation medium. (Right) After propagation, each receiver array element records dispersed signals from which the source array-receiver array transfer matrix is reconstructed.

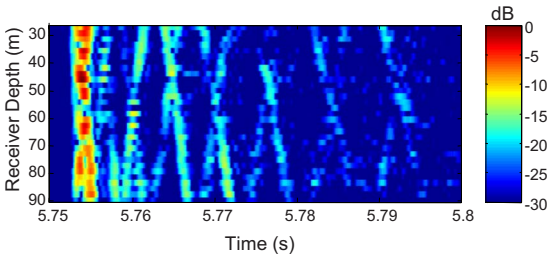


FIG. 3. (Color online) Spatiotemporal representation of the normalized acoustic intensity received at the receiver array for a pulsed signal emitted at a source array depth of 58 m. The data are from JD 105. The signal-to-noise ratio is on the order of 35 dB. The color bar is in dB.

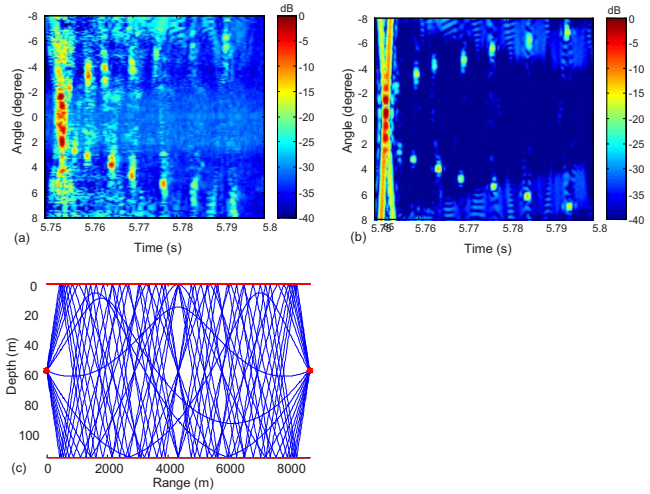


FIG. 4. (Color online) (a) Normalized intensity of the time-delay beamformer applied to the experimental pressure field recorded on the receiver array and shown in Fig. 3. (b) Normalized intensity of the time-delay beamformer performed on simulated data computed with a ray code for the same experimental configuration. (c) Eigenray paths between the source at 58 m depth and the receiver hydrophone at 60 m depth. Most of the rays are reflected by the ocean surface at least once.

Our goal was to evaluate the ratio between the coherent and incoherent intensities over multiply reflected echoes, as a way to characterize surface-height motion. The coherent and incoherent intensities are defined as

$$I_{\mathrm{coh}}(t) = \left[\frac{1}{P_{\mathrm{tot}}} \sum_{p=1}^{P_{\mathrm{tot}}} h_p(t) \right]^2, \tag{1}$$

$$I_{\mathrm{inc}}(t) = \frac{1}{P_{\mathrm{tot}}} \sum_{p=1}^{P_{\mathrm{tot}}} h_p^2(t), \tag{2}$$

where $h_p(t)$ is the signal received at a single hydrophone due to the pth transmission. Here P_{tot} is the total number of transmissions over which the average is performed. Note that some authors refer to the incoherent intensity in Eq. (2) as the total intensity (Ishimaru, 1978; Sheng, 1995). The coherent and incoherent definitions used here correspond to the type of averaging that is performed, i.e., the coherent intensity is defined as the intensity of the received signal coher-

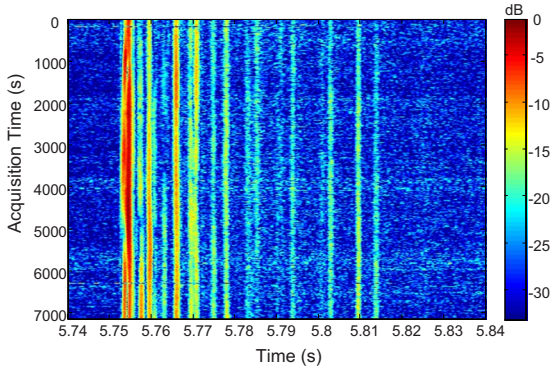


FIG. 5. (Color online) Normalized intensity of the impulse response for the signal transmitted by the 58-m-deep source and received by the 60-m-deep hydrophone during a nearly 2-h interval on JD 103. The pulse repetition rate was $1/30$ s^{-1}. Acquisition time corresponding to each transmission is along the y-axis. The horizontal axis is time after transmission. The color bar is in dB.

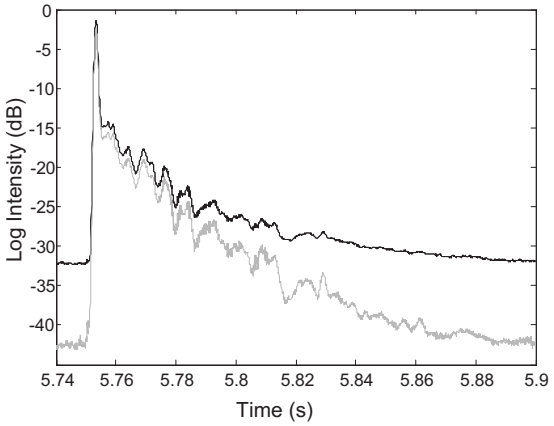


FIG. 6. Incoherent (black) and coherent (gray) intensities (in log scale) averaged over $M=29\times32=928$ source-receiver pairs. Here $P_{\text{tot}}=60$, which corresponds to a period of about 20 min on JD 103. Before coherent/incoherent averaging, each impulse response was normalized according to its maximum. Note that the coherent intensity falls off much faster than the incoherent intensity.

ently averaged over multiple pings, while the incoherent intensity corresponds to the average of the intensity. For the case of fast random perturbation of the ocean surface combined with a slowly varying medium such as the ocean volume (caused by internal waves, for example), a measurement of the coherent intensity associated with the ocean surface fluctuation only is (de Rosny and Roux, 2001)

$$I_{\text{coh}}(t)=\frac{1}{P_{\text{tot}}}\sum_{p=1}^{P_{\text{tot}}}h_p(t)h_{p+1}(t). \tag{3}$$

The CTIR is defined as

$$\text{CTIR}(t)=\left\langle\frac{I_{\text{coh}}(t)}{I_{\text{inc}}(t)}\right\rangle_M. \tag{4}$$

Here, the subscript M indicates the number of source–receiver pairs over which $I_{\text{coh}}(t)/I_{\text{inc}}(t)$ is averaged to obtain the CTIR. The coherent and incoherent intensities, calculated for $P_{\text{tot}}=60$, which corresponds to an integration of the wave height fluctuations over 20 min, and averaged over $M=928$ source-receiver pairs, are shown in Fig. 6. It is clear that $I_{\text{coh}}(t)$ falls off more quickly than $I_{\text{inc}}(t)$ which puts in evidence the effect of surface scattering on the coherent field.

The CTIR has been used over the last few years in diffusive reverberant acoustic wave spectroscopy (DRAWS) to measure the scattering cross section of small objects in motion in reverberant cavities (de Rosny and Roux, 2001; de Rosny *et al.*, 2005). In a reverberant cavity, the coherent intensity is due to reflections from the cavity boundaries, which do not change from transmission to transmission. When scatterers are present, the signal is different from one transmission to the next, and the CTIR quantifies the amount of coherent intensity that becomes incoherent because of the motion of the scatterer. DRAWS has been applied successfully at scales ranging from room acoustics at 1 kHz to ultrasonic cavities at 1 MHz (Conti *et al.*, 2004, 2006a).

For an infinite number of transmissions ($P_{\text{tot}}\to\infty$), the asymptotic expression for the CTIR classically decreases exponentially as $\text{CTIR}(t)=\exp(-t/\tau)$ (Ishimaru, 1978), where τ is defined as a scattering mean free time, from which the total scattering cross section of the scatterer can be determined. For a finite number of transmissions, it was shown in Appendix A of de Rosny and Roux, 2001 that the CTIR must include a second term that corresponds to the variance of P_{tot} random variables:

$$\widetilde{\text{CTIR}}(t)=\exp(-t/\tau)+\frac{1}{P_{\text{tot}}}(1-\exp(-t/\tau)). \tag{5}$$

The first exponential in Eq. (5) corresponds to the scattering effect of the scatterers, while the second term represents the residual variance of the scattered field. Indeed, for times much greater than the mean free time, i.e., $t\gg\tau$, the signal $h_p(t)$ is dominated by incoherent scattering from the moving objects rather than the coherent echoes from the cavity interfaces. In the following, we will alternatively use the *asymptotic* and *finite* notations $\text{CTIR}(t)$ and $\widetilde{\text{CTIR}}(t)$, respectively, for infinite and finite numbers of transmissions, respectively.

When there are no scatterers in the cavity, it has been demonstrated experimentally that the CTIR is affected, for example, by random motion of the cavity walls. As long as the ratio of volume to surface area of the cavity is large with respect to the acoustic wavelength, Eq. (5) remains valid and the CTIR still decreases exponentially (Conti *et al.*, 2006b; Conti, 2005). In this case, the mean free time τ is related to the rms displacement of the cavity walls.

However, the physical properties of an acoustic waveguide and a cavity are very different. In a cavity, reflections by the boundaries randomize the pressure field very quickly, in the sense that after a short transition time, a receiver at any location in the cavity will record a nearly continuous sequence of coherent acoustic echoes that have investigated the whole cavity volume. We classically describe this property of nontrivial cavities as ergodicity [see recent review by Tanner and Sondergaard (2007)]. As a result, the DRAWS technique performs well for cavities with only one, or at most a few, source-receiver pairs. The problem is very different in a waveguide where the pressure field is only bounded by two interfaces, above and below. Indeed, the acoustic field is no longer trapped in the volume of the cavity but can leak through the channel axis parallel to the waveguide interfaces. Going back to Figs. 3 and 5, we observe that the point-to-point signal in the waveguide is composed of a few arrivals due to interface reflections, surrounded by intervals during which there are no echoes. The use of the CTIR as a measure of the surface scattering mean free time is then problematic, since Eq. (3) can abruptly drop to zero, for time intervals where no coherent echoes are present. However, one way to mitigate the problem is to average the CTIR over a large number of sources and receivers. In so doing, the gaps between the coherent echoes are filled in when the CTIR is averaged [Eq. (4)] over a number of source-receiver pairs. Ultimately, the coherent intensity averaged over all of the source-receiver pairs shows no time discontinuity (see Fig. 6), which means that the temporal ergodicity of the cavity has been replaced by spatial diversity in the waveguide.

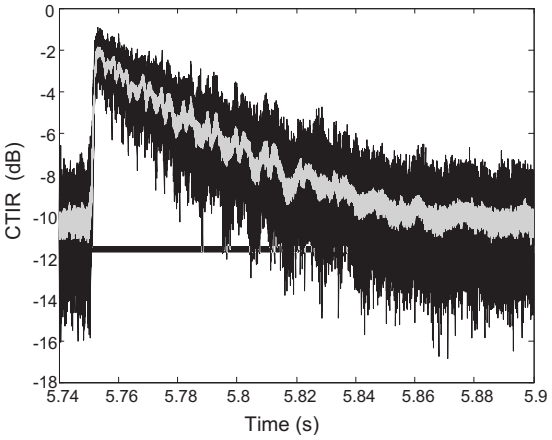


FIG. 7. Collections of 100 CTIRs measured on JD 103 for $P_{tot}=10$, and random selections of M source-receiver pairs among the 29-element source and 32-element receiver arrays: (a) $M=30$, black lines; (b) $M=700$, gray lines.

The benefit of spatial averaging is demonstrated in Fig. 7, which compares collections of CTIRs calculated for two different numbers of source-receiver pairs. A lower variance estimate of the CTIR is obtained by averaging over 700 source-receiver pairs. This confirms that spatial averaging over sources and receivers that are distributed over the water column provides arrival-time diversity, which reduces the variance of the CTIR.

Note that in Fig. 7 (and also in Figs. 8 and 9), the CTIR value at $t=t_0$ (where $t_0=5.753$ s corresponds to the travel time of the first arrival) is not equal to 1.0, which means that the coherent intensity differs from the incoherent intensity even for the direct arrival. One possible explanation is a small ($\sim 1/8F_c$) jitter that is present in the source timing. It is also possible that random fluctuations in the ocean volume along the direct path were responsible for the drop in coherence.

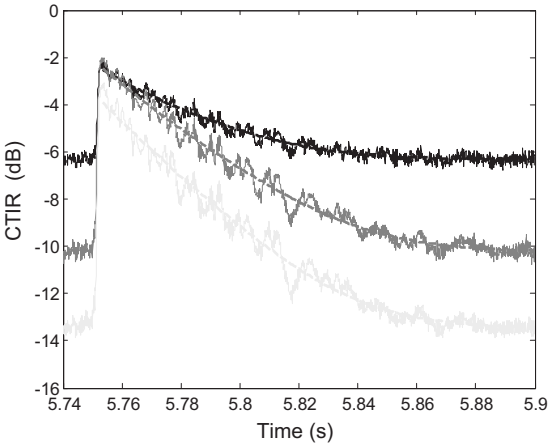


FIG. 8. CTIR calculated for the same data on JD 103 using $M=928$ source-receiver pairs and different numbers of transmissions P_{tot}: $P_{tot}=20$ (light gray, bottom curve), $P_{tot}=10$ (gray, middle curve), and $P_{tot}=5$ (black, top curve). The dashed curves correspond to the best CTIR fit obtained using the Kirchhoff approximation [Eq. (12)] and rms surface heights of $h=0.18$ m, $h=0.195$ m, and $h=0.177$ m, respectively. Note that the rms surface heights agree to within about 5%, regardless of P_{tot} and despite the significant effect of P_{tot} on the CTIR.

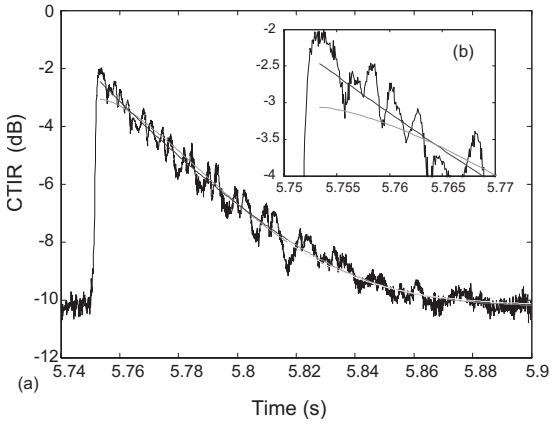


FIG. 9. (a) CTIR measurement (solid black line) compared to CTIR predictions based upon the Kirchhoff approximation (light gray) and the SSA approach (dark gray). The CTIR was calculated with $P_{tot}=10$ and $M=928$ source-receiver pairs. (b) Expanded view of the CTIR for small δt showing better agreement between the experimental data and the SSA at shallow grazing angles. Applying the Kirchhoff approximation gives $h=0.179$ m [corresponding to $U=5.63$ m/s using Eqs. (12) and (13)], while the SSA approximation Eqs. (18)–(20) yields $U=5.87$ m/s and $h=0.184$ m.

III. THEORY RELATING THE CTIR TO SURFACE SCATTERING IN AN ACOUSTIC WAVEGUIDE

Repetitive measurements of the transfer matrix between the source and receiver arrays provide a comparison between CTIR measurements and both wind speed and significant wave height over a few days. Significant wave heights were extracted from continuous recordings on a wave-rider buoy, and wind speed was measured on the research vessel Alliance. The goal of this section is to determine how the CTIR is related to the surface conditions and to derive the relationship between the measured CTIR and the significant wave height and/or wind speed.

Medwin and Clay (1994) showed that in the Kirchhoff approximation, the coherent reflection coefficient for an incident plane wave at frequency ω forward scattered by a surface having a Gaussian height probability distribution function is

$$\mathrm{CSRC}_{\mathrm{Kirchhoff}}(\theta,\omega)=\Re\,\exp(-2k^2h^2\sin^2\theta), \tag{6}$$

where $k=\omega/c$ is the acoustic wave number (ω is angular frequency and c is the speed of sound), h^2 is the mean square wave height of the ocean surface, and θ is the grazing angle. For the air-water interface, $\Re=-1$. Equation (6) is based on the assumption that a plane wave reflected by a rough surface will experience a phase shift of $2k\delta z\sin\theta$, where δz is the height of the surface at the specular point relative to the mean surface level. The coherent reflection coefficient in Eq. (6) is obtained by taking the mean value of $\exp(i2k\delta z\sin\theta)$ over the surface-wave height.

Following the Kirchhoff approximation, the coherent *intensity* surface reflection coefficient (CISRC) was defined by Thorsos (1990) as

$$\mathrm{CISRC}_{\mathrm{Kirchhoff}}(\theta,\omega)=\exp(-4h^2k^2\sin^2\theta). \tag{7}$$

The CISRC assumes a plane wave incident on a rough surface at grazing angle θ and it characterizes the average intensity of the wave reflected at the same grazing angle. The

average here corresponds to an ensemble average over many realizations of the rough surface. When there are N surface reflections, as for longer-range propagation in shallow water, the CISRC must be applied N times:

$$\mathrm{CISRC}_{\mathrm{Kirchhoff}}(\theta,\omega)=\exp(-4Nh^2k^2\sin^2\theta). \tag{8}$$

For the experimental results presented here, the quasi-uniform sound speed profile leads to the absence of refraction in the waveguide, which means that the incident and reflected angles are conserved through the waveguide propagation.

As an acoustic signal travels through a shallow, constant-sound speed channel, in addition to N scattering events at the ocean surface, it also undergoes Q reflections at the bottom. The Q bottom interactions, the volume attenuation, and the spreading loss at range r can be combined into an intensity loss $A(Q,\omega,\theta,r)$, which is in addition to the intensity loss expressed in Eq. (8). However, the intensity loss factor $A(Q,\omega,\theta,r)$ that results from propagation through the channel will be incorporated into the incoherent intensity, which has the role of a normalization factor in the CTIR. Therefore, the CTIR reduces to the averaged intensity loss due to surface scattering only, as expressed in Eq. (8).

However, Eq. (8) was derived for a plane wave propagation in the waveguide (at grazing angle θ and frequency ω), and it is no longer valid in the case of an omnidirectional source and receiver (Fig. 1) for which all angles are mixed up in the frequency domain. Fortunately, things can be sorted out in the time domain. Indeed, the relationship between grazing angle and travel time can be easily determined in a range- and depth-independent sound speed environment, such as a shallow-water channel. An eigenray at grazing angle θ that arrives at time t at the end of the waveguide satisfies the relationship

$$\sin^2(\theta)=\delta t\frac{\delta t+2t_0}{(\delta t+t_0)^2}\approx 2\frac{\delta t}{t_0}, \tag{9}$$

where $t_0=5.753$ s is the travel time of the direct path between the source and the receiver in the waveguide, and where $\delta t=t-t_0$. The approximation in Eq. (9) requires that $\delta t\ll t_0$, which is normally the case in shallow water at longer ranges [Fig. 4(a)]. Also, the number of surface reflections, N, is time dependent and increases with travel time, from $N=2$ at $\delta t=0.005$ s to $N=7$ at $\delta t=0.12$ s in our experimental configuration. More generally, for a shallow-water channel with depth D and range L, the relationship between time and the number of surface bounces N is

$$N(\delta t)=\frac{L}{2D}\sqrt{\frac{\delta t}{\frac{t_0}{2}-\delta t}}\approx\frac{L}{\sqrt{2}D}\sqrt{\frac{\delta t}{t_0}}. \tag{10}$$

Substituting Eqs. (9) and (10) into Eq. (8), and taking the CTIR to be equal to the average intensity loss due to surface scattering only, after frequency summation over a large bandwidth, the time-dependent CTIR can be written as

$$\mathrm{CTIR}_{\mathrm{Kirchhoff}}(\delta t)\approx\exp\left(-\frac{4\sqrt{2}h^2\omega_c^2(\delta t)^{3/2}}{Dc\sqrt{t_0}}\right), \tag{11}$$

where $\omega_c=2\pi F_c$ is the central frequency of the pulse. Note that the transition from continuous wave to time domain in Eqs. (8)–(11) is not a Fourier transform, since no phase information is present in the CTIR.

Actually, Eq. (11) corresponds to the *asymptotic* CTIR for an infinite number of transmissions in the acoustic waveguide. In practice though, noise must be added to Eq. (11), since the CTIR has no physical meaning when no coherent signal is present, for example, at $t_0+\delta t>5.82$ s in Fig. 5. The floor of the CTIR corresponds to the variance of the coherent to incoherent ratio of the random ocean noise recorded on the array, and it only depends on the number of transmissions, P_{tot}, over which the average is performed. Writing Eq. (11) in the form of Eq. (5), the expression for the *finite* CTIR is then

$$\widetilde{\mathrm{CTIR}}_{\mathrm{Kirchhoff}}(\delta t)\approx\exp\left[-\left(\frac{\delta t}{\tau}\right)^{3/2}\right]+\frac{1}{P_{\mathrm{tot}}}\left(1-\exp\left[-\left(\frac{\delta t}{\tau}\right)^{3/2}\right]\right) \tag{12}$$

with

$$\tau^{3/2}=\frac{Dc\sqrt{t_0}}{4\sqrt{2}h^2\omega_c^2}. \tag{13}$$

Figure 8 shows the CTIR calculated for the same data on JD 103, but using values of $P_{\mathrm{tot}}=5$ (black, upper curve), 10 (gray, middle curve), and 20 (light gray, lower curve). Clearly, the CTIR falls off more quickly with increasing P_{tot}. However, we show now that this dependence is properly reflected in Eq. (12), such that we can use the CTIR to predict surface-wave height quite well for different values of P_{tot}. Equation (12) has been used to find the values of rms wave height h that provide the CTIR that is the best match with that calculated from the data. The dashed lines in Fig. 8 indicate the CTIR calculated using Eq. (12) and the values of $h=0.18$, 0.195, and 0.177 m, and $P_{\mathrm{tot}}=20$, 10, and 5, respectively. Although these values for h are not identical, they only vary by about 5%, indicating that different values of P_{tot} can be used to compute the CTIR and to estimate the rms surface height without significantly affecting the result. The results in Fig. 8 confirm the dependence of the CTIR on the number of transmissions, P_{tot}, which must be taken into account when the CTIR is used to estimate the ocean surface condition.

The solid black line in Fig. 9 shows the CTIR calculated from the data on JD 103 for $P_{\mathrm{tot}}=10$, averaged over 928 source-receiver pairs. The light gray line is the best fit CTIR calculated using Eqs. (12) and (13). Although the best fit Kirchhoff CTIR matches the calculated CTIR over most of the curve, it does so at the expense of some disagreement at small δt. These times, which are shortly after the direct path arrival, correspond to paths which have interacted with the surface at very low grazing angles. In search of a better fit at low grazing angles, we considered the SSA to the CSRC as providing a more accurate reflection coefficient for shallow grazing angles (Thorsos and Broschat, 1995). Using the

same notations as in Eq. (6), the CSRC for an incident plane wave (at frequency ω and grazing angle θ) in the SSA becomes (Broschat, 1993):

$$\mathrm{CSRC}_{\mathrm{SSA}}(\theta,\omega) = \Re \exp(-2k^2h^2 \sin^2\theta) \times \left| 1 - 2k \sin\theta \int W(K)g(K,\theta,\omega)dK \right|, \tag{14}$$

where K is the one-dimensional (1D) surface-wave number, $W(K)$ is the 1D surface roughness spectrum, $g(K,\theta,\omega) = \sqrt{k^2-(K+k\cos\theta)^2}-k\sin\theta$, and $\Re=-1$, as in Eq. (6).

A commonly used ocean two-dimensional (2D) surface-height spectrum that represents a fully developed sea was published by Pierson and Moskowitz (1964). The Pierson–Moskowitz surface-height spectrum is completely determined by the wind speed. Given the geometry of our measurements, with coplanar sources and receivers, we can use a simplified version of the 2D Pierson–Moskowitz spectrum, which leads to the following 1D roughness spectrum (Broschat, 1993):

$$W(K) = (\alpha/(4|K|^3))\exp[-(\beta g^2)/(K^2U^4)]. \tag{15}$$

Here α and β are dimensionless constants of 8.1×10^{-3} and 0.74, respectively, $g=9.81$ m/s^2 is the gravitational acceleration constant of Earth, and U is the wind speed measured at 19.5 m above the mean ocean surface. The spectrum is normalized such that

$$\int W(K)dK = h^2. \tag{16}$$

When there are N surface reflections, as for longer-range propagation in shallow water, the asymptotic CISRC (at frequency ω and grazing angle θ) in the small-slope approximation is

$$\mathrm{CSRC}_{\mathrm{SSA}}(\theta,\omega) = \exp(-4Nh^2k^2 \sin^2\theta) \times \left| 1 - 2k \sin\theta \int W(K)g(K,\theta,\omega)dK \right|^{2N}. \tag{17}$$

Finally, using Eqs. (9) and (10) and accounting for the finite number of transmissions P_{tot} as in Eqs. (5) and (12), the expression for the time-dependent *finite* CTIR after frequency integration around ω_c (with $k_c=\omega_c/c$) is

$$\widetilde{\mathrm{CTIR}}_{\mathrm{SSA}}(\delta t) \approx \exp\left[-\left(\frac{\delta t}{\tau}\right)^{3/2}\right]F(\delta t) + \frac{1}{P_{\mathrm{tot}}}\left(1-\exp\left[-\left(\frac{\delta t}{\tau}\right)^{3/2}\right]F(\delta t)\right), \tag{18}$$

where

$$F(\delta t) \approx \left| 1 - k_c\sqrt{\frac{8\delta t}{t_0}} \int W(K)g(K,\delta t)dK \right|^{L/D\sqrt{2\delta t/t_0}} \tag{19}$$

and

$$g(K,\delta t) = \sqrt{k_c^2 - \left(K + \frac{k_c t_0}{t_0+\delta t}\right)^2} - 2k_c\frac{\delta t}{t_0}. \tag{20}$$

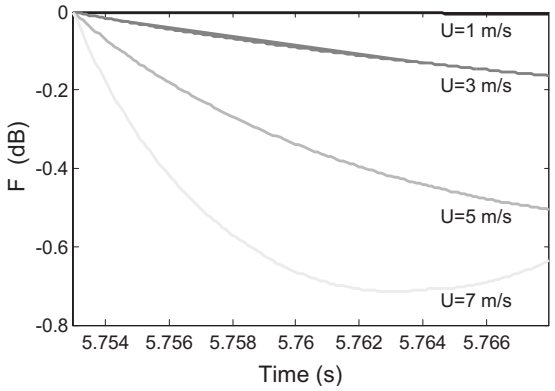


FIG. 10. Log representation of the short-time correction $F(\delta t)$ (corresponding to low grazing angle) for the time-dependent CTIR in the SSA approximation [Eq. (18)] and for different values of wind speed, U (ranging from $U=1$ m/s to $U=7$ m/s)

In Eq. (18), $F(\delta t)$ is a short-time correction factor for the time-dependent CTIR; in the same way, the second term in Eq. (17) is a low-angle correction to the frequency-dependent CISRC in the Kirchhoff approximation. As shown in Fig. 10, the factor $F(\delta t)$ is less than 1 for $\delta t>0$ but small. This serves to straighten the log representation of the CTIR at short times $0<(\delta t/\tau)\ll1$ as may be seen in Fig. 9(b), where the CTIR calculated using Eqs. (18)–(20) is compared to the time dependence in $-(\delta t/\tau)^{3/2}$ obtained from the Kirchhoff approximation [Eq. (12)].

Figure 9(b) shows that utilizing the SSA provides better agreement with the experimental data at early times. The best match using Kirchhoff approximation gives $h=0.179$ m, from which we calculate $U=5.63$ m/s using Eqs. (15) and (16), while the best match using the SSA approximation gives $U=5.87$ m/s and $h=0.184$ m. These values of rms surface-wave height h and wind speed are very similar. Indeed, the main surface scattering contribution seen in the experimental data for 10 ms$<\delta t<$70 ms comes from grazing angles $3.5°<\theta<9°$ and wind speeds $U>5$ m/s for which the SSA and the Kirchhoff approximation give similar results (Williams *et al.*, 2004).

IV. DISCUSSION

Equations (12) and (18) show that in an acoustic waveguide, the CTIR does not decrease exponentially with time, as was the case for multiple scattering theory in an acoustic cavity [Eq. (5)]. Indeed, the CTIR calculated using both the Kirchhoff and small-slope approximations for surface scattering varies exponentially with $-(\delta t/\tau)^{3/2}$ over most of the arrival times. In analogy with multiple scattering theory, we deduce an equivalent to the concept of mean free time τ [see Eq. (13)] or equivalently an e-folding time, for an acoustic waveguide with random surface scattering. In an acoustic waveguide, the scattering mean free time τ is a measure of the time interval after the direct echo after which the reverberated pressure field becomes largely incoherent. From Fig. 9 and Eq. (13), we obtain $h=0.18$ m and $\tau\sim35$ ms. This value can be compared to the attenuation mean free time ($\sim$60 ms) evident in the incoherent intensity (see Fig. 6). In our shallow-water configuration, τ is a measure of the time

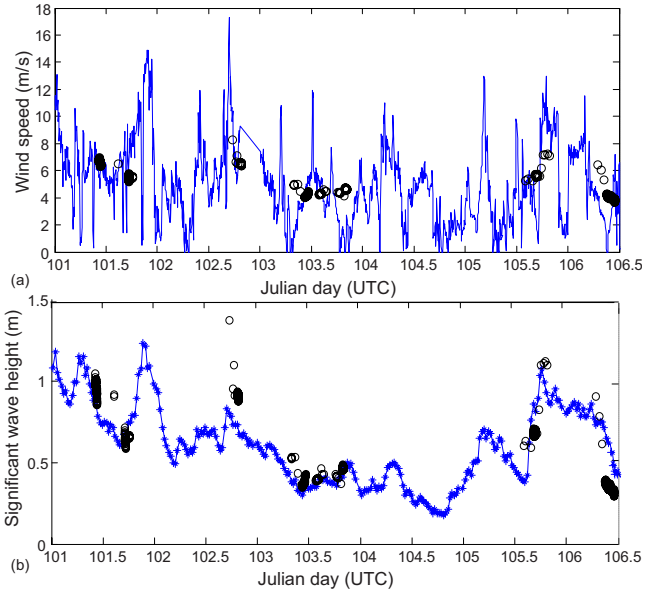


FIG. 11. (Color online) Wind speed measurements [solid line in (a)] and significant wave height H_S [dashed line in (b)] measured using a wave-rider buoy for JDs 101–106 ($H_S=4h$, with h the rms surface wave height). The open circles are wind speed and significant wave height obtained by fitting Eq. (18) to the CTIR calculated from the acoustic measurements.

after the direct arrival when the received signal becomes largely incoherent from transmission to transmission (see Fig. 5, for example). The measure of the scattering mean free time could be important when dealing with coherent underwater acoustic communications or time-reversal in shallow-water waveguides.

We have shown that Eqs. (12) and (18) allow us to invert for surface scattering properties from the CTIR measurements performed between the two vertical source-receiver arrays. In the Kirchoff approximation, the CTIR only depends on the rms surface height h, while the wind speed U is also obtained using the SSA approximation and the surface-wave height spectrum.

Carrying out this inversion for all of the acoustic data, Fig. 11 compares the wind speed and significant wave height ($H_S=4h$, with h the rms surface-wave height) estimated from the acoustic data by finding the best fit CTIR with measurements of these parameters for JDs 101–106. The wave height measurements are obtained from wave-rider-buoy data. It is clear that the CTIR-derived significant wave height estimates are in good agreement with the measured values. The wind speed is much more volatile than the wave, and of course, ocean surface waves are a response to the wind blowing over the ocean surface. The CTIR is directly affected by surface roughness, so it is not surprising that it is better correlated with surface condition wind speed. The largest discrepancy occurs for the significant wave height between JD 102.5 and JD 103. Note, however, that the higher wave height estimates do correspond to a large peak in measured wind speed that only yields a modest peak in significant wave height.

In conclusion, we have shown that the CTIR is a measure of the interference effects which cause propagating waves to fluctuate from transmission to transmission. In the case of an acoustic waveguide, the fluctuations can be caused by time-varying volume or surface scattering. Surface scattering was dominant in the experimental configuration presented here. Also, the essentially isovelocity sound speed profile seen here in April indicates that minimal volume variation can be expected during the averaging times we employed. However, we note that during the summer, the presence of a strong thermoclyne could induce internal waves that would cause significant time-varying volume refraction on a short time scale. In this case, both volume refraction and the surface scattering effects would be measured simultaneously by the CTIR. The two contributions could be sorted out if refracted or reflected rays are identified and considered separately in the CTIR calculation, as shown recently using source-receiver arrays and a double-beamforming algorithm (Roux *et al.*, 2008). Future work will combine the CTIR and eigenray identification to separately invert for surface scattering and volume scattering effects in an acoustic waveguide.

ACKNOWLEDGMENTS

The authors would like to thank W. A. Kuperman, W. H. Hodgkiss, Mark Stevenson, and the NATO Saclant Center for access to the FAF03 data.

Broschat, S. L. (**1993**). "The small slope approximation reflection coefficient for scattering from a 'Pierson-Moskowitz' sea surface," IEEE Trans. Geosci. Remote Sens. **31**, 1112–1114.

Conti, S. G. (**2005**), "Characterization of scatterers in reverberating cavities and applications," Ph.D. thesis, University of Paris 6, Paris, France.

Conti, S. G., Roux, P., de Rosny, J., and Demer, D. A. (**2006b**). "Characterization of scatterer motion in a reverberant media," J. Acoust. Soc. Am. **119**, 769–776.

Conti, S. G., Roux, P., Demer, D. A., and de Rosny, J. (**2004**). "Measurement of the scattering and absorption cross-sections of the human body," Appl. Phys. Lett. **84**, 819–821.

Conti, S. G., Roux, P., Fauvel, C., Maurer, B. D., and Demer, D. A. (**2006a**). "Acoustical monitoring of fish density, behavior, and growth rate in a tank," Aquaculture **251**, 314–323.

Dahl, P. H. (**2004**). "Forward scattering from the sea surface and the van Cittert-Zernike theorem," J. Acoust. Soc. Am. **115**, 589–599.

de Rosny, J., and Roux, P. (**2001**). "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597.

de Rosny, J., Tourin, A., Derode, A., Roux, P., and Fink, M. (**2005**). "Weak localization and time reversal of ultrasound in a rotational flow," Phys. Rev. Lett. **95**, 074301.

Ishimaru, A. (**1978**). *Wave Propagation and Scattering in Random Media* (Academic, New York), Chap. 14.

Lewis, J. K., Rudzinsky, J., Rajan, S., Stein, P. J., and Vandiver, A. (**2005**). "Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones," J. Acoust. Soc. Am. **117**, 3539–3554.

Lynch, J. F., Spindell, R. C., Chiu, C. S., Jimmer, J. H., and Birdsall, T. G. (**1987**). "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.

Medwin, H., and Clay, C. S. (**1994**). *Fundamentals of Acoustical Oceanography* (Academic, San Diego).

Miller, J. H., Lynch, J. F., and Chiu, C. S. (**1989**). "Estimation of sea surface spectra using acoustic tomography," J. Acoust. Soc. Am. **86**, 326–345.

Miller, J. H., Lynch, J. F., Chiu, C.-S., Westreich, E. L., Hippenstiel, R., and Chaulk, E. (**1993**). "Acoustic measurement of surface gravity wave spectra in Monterey Bay using mode travel-time fluctuations," J. Acoust. Soc. Am. **94**, 954–974.

Ogilvy, J. A. (**1987**). "Wave scattering from rough surfaces," Rep. Prog. Phys. **50**, 1553–1608.

Pierson, W. J., Jr., and Moskowitz, L. (**1964**). "A proposed spectral form for fully developed wind seas based on the similarity theory of S.A. Kitaigorodskii," J. Geophys. Res. **69**, 5181–5190.

Roux, P., Cornuelle, B. D., Kcuperman, W. A., and Hodgkiss, W. S. (**2008**). "The structure of ray-like arrivals in a shallow water waveguide," J. Acoust. Soc. Am. **124**, 3430–3439.
Roux, P., Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., and Stevenson, M. (**2004**). "A non-reciprocal implementation of time reversal in the ocean," J. Acoust. Soc. Am. **116**, 1009–1015.
Sheng, P. (**1995**). *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego), Chaps. 3 and 4.
Tanner, G., and Sondergaard, N. (**2007**). "Wave chaos in acoustics and elasticity," J. Phys. A **40**, R443–R509.
Thorsos, E. I. (**1988**). "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **83**, 78–92.
Thorsos, E. I. (**1990**). "Acoustic scattering from a "Pierson-Moskowitz" sea surface," J. Acoust. Soc. Am. **88**, 335–349.
Thorsos, E. I., and Broschat, S. L. (**1995**). "An investigation of the small slope approximation for scattering from rough surfaces. Part I. Theory," J. Acoust. Soc. Am. **97**, 2082–2093.
Thorsos, E. I., and Jackson, D. R. (**1989**). "The validity of the perturbation approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **86**, 261–277.
Williams, K. L., Thorsos, E. I., and Elam, W. T. (**2004**). "Examination of coherent surface reflection coefficient (CSRC) approximations in shallow water propagation," J. Acoust. Soc. Am. **116**, 1975–1984.
Yang, C. C., Fennemore, G. C., and McDaniel, S. T. (**1992**). "Scintillation index of the acoustic field forward scattered by a rough surface for two- and three-dimensional scattering geometries," J. Acoust. Soc. Am. **91**, 1960–1966.

Asymptotic evaluation of the pulse train radiated by an angled beam and fluid coupled rectangular ultrasonic transducer

D. D. Zakharov
LMP, Universite Bordeaux 1, 351 cours de la Liberation, 33405 Talence, France

L. Ju. Fradkin
FESBE, London South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom

(Received 30 November 2008; revised 12 August 2009; accepted 22 December 2009)

The near field underneath the ultrasonic probe fluid coupled to an isotropic solid is approximated in the frequency domain by a closed form asymptotic solution. The approximation is based on the problem decomposition and uses the stationary phase method evaluating the response to an equivalent surface source. This results in a sum of contributions, each dominating in a specific geometrical region, the main beam or a side lobe. The transitional zones are also described. The pulse trains are computed using the harmonic synthesis and compared with the results obtained by direct calculation of Fourier integrals. It is shown that the asymptotic approach permits us to elucidate the physics of problem and leads to a numerical algorithm which is about 10^4 times faster than the direct computations. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3292997]

PACS number(s): 43.35.Cg, 43.20.Dk, 43.35.Zc, 43.20.Px [ADP] Pages: 1267–1278

I. INTRODUCTION

It is well known that two types of ultrasonic probes commonly used in non-destructive evaluation (NDE) are the normal transducer and the angled beam probe,[1] operating typically from 1 to 10 MHz. The normal transducer is a piezoelectric crystal placed directly onto the inspection surface and it generates a compression beam normal to this surface. The angle beam probe[1] consists of a piezoelectric crystal mounted on a plastic wedge and it generates a beam of shear waves propagating at a specific angle to the surface. In both cases, the contact with the inspection solid is through a thin layer of low-viscous liquid couplant, which permits us modeling the equivalent surface source[2] with neglected shear stresses. For the transducer's far field, a moving point source[3–5] can be used to model such equivalent surface sources. A line source or strip source can be used[6–8] for modeling a normal rectangular probe. Both the far field[9] and near field can also be satisfactory modeled,[10] e.g., using the direct computations or the asymptotic approach.[11–13] A direct computation, performed using finite element method (FEM) (Ref. 14) or finite-differences,[15] may produce satisfactory simulations, but the computational effort at the high frequency and/or large distance can be a drawback. As shown in Ref. 16, the head waves are prominent features of the wave field, radiated by a small contact transducer and probe with steep beam angle, while the Rayleigh waves always dominate at the surface. Models that neglect the mode conversions at the free surface fail to predict this field component.[12,17,18] To describe the full wave field underneath the transducer accurately and at any range, the exact elastodynamic solution is required. Such solution could be obtained by employing the point-source superposition. However, in a three dimensional geometry this results in a triple integral over the time and aperture and requires an enormous computational time. This time increases with increasing pulse width since the spatial discretization depends on the temporal sampling rate, and in view of the Shannon sampling theorem, the latter must be relatively high. Although combining the approximations with smart programming may considerably decrease the computational effort,[19] the evaluation may still suffer from invalid approximation. For some source shapes, e.g., circular, annular, and rectangular sources,[20–25] the spatial (Rayleigh) integral can be evaluated semi-analytically, e.g., using the contour integration, non-moving line segment source, and the Cagniard–de Hoop method.[5,16] This approach is less computationally extensive but still remains time consuming. In this paper we apply a different technique, which has first been used to describe the field of the directly coupled normal probe.[11–13,26] We solve the problem in several steps. First, we evaluate the magnitude of the surface load from a *local* problem assuming that the couplant layer is thin and solving in the frequency domain a simple problem of the contact of two infinite solids through a fluid layer. Since in practice we never know the exact shape and size of a contact spot between the couplant and inspection solid, neglecting finite size looks reasonable. We thus estimate stresses on the liquid-solid interface. We then follow the traditional formulation and model the probe by an equivalent surface source—a contact spot—which generates the above stresses, assuming them to fall to zero outside the spot in a discontinuous manner—even though the real boundary conditions may be more involved. Second, we adapt the two-tier asymptotic approach to evaluate the radiation field in the frequency domain. To this end we use the solution of Lamb's problem approximated by its leading term at high frequency (like the far field) and apply the stationary phase method to evaluate the convolution integrals over equivalent contact spot considering the first Fresnel zone underneath the transducer. At the third step, we apply the time-harmonic synthesis to the one obtained above and simulate the pulse trains propagating in this region. In principle, assuming that the discontinuous change in stresses is

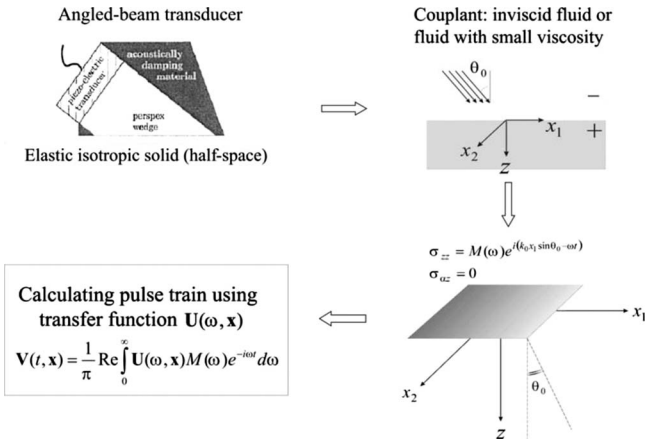


FIG. 1. The problem decomposition.

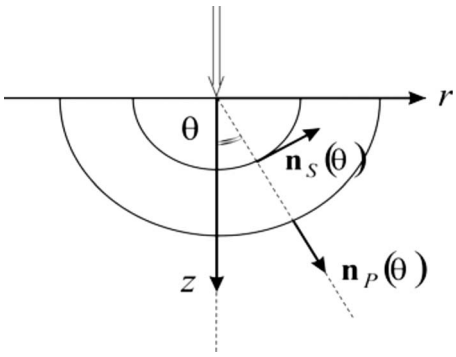


FIG. 2. The wave fronts and polarization vectors.

not necessary, the approach is applicable to probes of arbitrary apodization. Keeping this in mind, the model of the contact spot may be improved, e.g., by utilizing the measured pressure values underneath the probe.[16] Of course, if the stress decay near the boundary of the contact spot is continuous, the contribution of the side lobes is less perceptible than the discontinuous model would imply.

This paper is organized as follows. In Sec. II we present our decomposition concept using steps from 1 to 3. Then we consider the convolution integral, apply the stationary phase method, and obtain the main beam, edge and corner waves (Sec. III), and penumbra zones (Secs. IV–VI) in the frequency domain. The numerical simulation of pulse trains and their comparison with the direct evaluation of field radiated by the same contact spot are presented in Secs. VII and VIII, followed by some conclusive remarks.

II. PROBLEM DECOMPOSITION CONCEPT

The angled beam ultrasonic probe placed on the surface of the inspection solid (see Fig. 1, top left) may be modeled using the following problem decomposition. First, we study the wave generated by the piezoelectric element and incident on the transducer-fluid couplant-inspection surface on the scale of the fluid couplant. The couplant layer is assumed to be thin compared to the characteristic transducer dimensions and of infinite extent. We solve this problem to arrive at a transfer function between plastic wedge, fluid couplant, and inspection solid (see Fig. 1, top right) and to obtain the stress field on the inspection surface. We prescribe this stress distribution on the contact spot on the inspection surface which we treat as an equivalent source, whose stress falls to zero outside the spot in a discontinuous manner (marked in Fig. 1 with "+"). The model may be refined by taking into account the continuous distribution but such correction does not change the approach too much. The above distribution is a traveling wave, whose magnitude depends on the frequency, incidence angle, material parameters, and thickness of the couplant. It can be obtained in a closed form or using low frequency asymptotics (see, e.g., Ref. 27). Second, we use the so-called "two-tier high frequency asymptotic approach"[11] to model the near radiating field of the equivalent contact spot. Third, we use the time convolution of the given pulse with the above results to model the pulse trains arriving at a given point of observation. Let us overview the two-tier approach.

1) *The first tier*: The high frequency asymptotics of the time-harmonic Green function (Lamb problem). Since our interest is focused on the radiating field, that is, a few wavelengths away from the surface, in the first tier we arrive at the far field asymptotics of the Lamb's problem (the Green's function for the isotropic half space). We use it to approximate the convolution integral, interpreted as a superposition of fields radiated by point Huygens's sources covering the surface of the contact spot. The leading components of the displacement vector (the factor $e^{-i\omega t}$ is omitted) are presented in Refs. 18 and 28. In the local Cartesian base $\mathbf{x}=\mathbf{i}_\alpha x_\alpha$ ($\alpha=1,2,3$) and associated cylindrical coordinates r,θ,z with the origin at the point-source (Fig. 2), the displacement vector is a sum of P- and S-wave components

$$\mathbf{U}^{P,S}(\mathbf{x}) = -\frac{1}{2\pi\mu}\frac{e^{ik_{P,S}R}}{R}A_{P,S}(\mathbf{x})\mathbf{n}_{P,S}(\mathbf{x}) + O(k_{P,S}R)^{-2}, \quad k_{P,S}z = 2\pi z/\lambda_{P,S} \gg 1, \tag{1}$$

where μ is the shear modulus and

$$R(\mathbf{x}) = \sqrt{r^2+z^2}, \quad r(\mathbf{x}) = \sqrt{x_1^2+x_2^2}, \quad z = x_3, \quad \gamma = c_S/c_P < 1, \quad k_{P,S} = \omega/c_{P,S}, \quad 0 < \theta < \pi/2,$$

$$A_P(\mathbf{x}) = \frac{2\sin^2\theta - \gamma^{-2}}{R_P(\sin\theta)}\cos\theta, \quad R_P(\eta) = (2\eta^2-\gamma^{-2})^2 + 4\eta^2\sqrt{1-\eta^2}\sqrt{\gamma^{-2}-\eta^2}, \quad \mathbf{n}_P(\mathbf{x}) = \sin\theta\mathbf{e}_r + \cos\theta\mathbf{e}_z,$$

$$A_S(\mathbf{x}) = \frac{\sin 2\theta\sqrt{\gamma^2-\sin^2\theta}}{R_S(\sin\theta)}, \quad R_S(\eta) = (2\eta^2-1)^2 - 4\eta^2\sqrt{\eta^2-1}\sqrt{\eta^2-\gamma^2}, \quad \mathbf{n}_S(\mathbf{x}) = \mathbf{e}_r\cos\theta - \mathbf{e}_z\sin\theta.$$

At the angle $\theta \approx \pi/2$ these leading terms vanish, and the second order contribution must be taken into account, but away from this angle, the approximation is accurate and the *head wave* $\mathbf{U}^H = O(k_S R)^{-2}$ and can be neglected.[28]

2) *The second tier*: High frequency asymptotics of the radiating near field. Assume that the observation point lies in

the first Fresnel zone $z/k_{P,S}[\min(a_1,a_2)]^2 \ll 1$ underneath the equivalent contact spot $\Omega \equiv [-a_1,a_1]\times[-a_2,a_2]$ with given stresses $\sigma_{zz}=M(\omega)e^{ik_0 x_1 \sin\theta_0}$, $\sigma_{\alpha z}=0$; so the incident wave is supposed to fall along x_1. Using the superposition of point-sources and the first tier from above, we express the leading term of the displacement field normalized by $M(\omega)$ in the form of the two dimensional (2D) integral of rapidly oscillating function where the asterisk labels *P*- and *S*-wave components, $\mathbf{x}$ is the observation point, and c_0 is the sound speed in fluid:

$$\mathbf{U}^*(\mathbf{x}) \cong -\frac{1}{2\pi\mu}\int\int_{\Omega} \mathbf{f}_*(\mathbf{x},\mathbf{x}')e^{ik_*\Psi_*(\mathbf{x},\mathbf{x}')}dx_1'dx_2',$$

$$\mathbf{f}_*(\mathbf{x},\mathbf{x}') = \frac{A_*(\mathbf{x}-\mathbf{x}')\mathbf{n}_*(\mathbf{x}-\mathbf{x}')}{R(\mathbf{x}-\mathbf{x}')}, \tag{2}$$

$$\Psi_*(\mathbf{x},\mathbf{x}') = R(\mathbf{x}-\mathbf{x}') + x_1' \sin\theta_*, \quad \sin\theta_* = \frac{c_*}{c_0}\sin\theta_0,$$

$$R(\mathbf{x}-\mathbf{x}') = \sqrt{(x_1'-x_1)^2+(x_2'-x_2)^2+z^2},$$

$$\mathbf{x}' = (x_1',x_2',0) \in \Omega. \tag{3}$$

The high frequency asymptotics of this near field is our "second tier."[11]

III. STATIONARY PHASE METHOD: MAIN BEAM, EDGE, AND CORNER WAVES

Consider the double integral (2) at $z>0$ and obtain the derivatives of its phase function (δ_α^1 is a Kronecker delta with $\alpha=1,2$)

$$\Psi_\alpha \equiv \partial_{x_\alpha'}\Psi_* = (x_\alpha'-x_\alpha)R^{-1} + \delta_\alpha^1 \sin\theta_*,$$

$$\Psi_{\alpha\alpha} \equiv \partial_{x_\alpha'}\Psi_\alpha = R^{-1} - (x_\alpha'-x_\alpha)^2R^{-3},$$

$$\Psi_{12} \equiv \partial^2_{x_1'x_2'}\Psi_* = -(x_1'-x_1)(x_2'-x_2)R^{-3},$$

$$\det\Psi'' \equiv \Psi_{11}\Psi_{22}-\Psi_{12}^2 = z^2R^{-4} > 0.$$

It is well known that, apart from singularities of the amplitude function $\mathbf{f}_*$, the main contributions to the integral (2) come from the *stationary points* of the phase functions, i.e., the points where $\Psi_1=\Psi_2=0$ (when isolated, these points are called the *critical points of the first kind*[29,30]), and from the various types of *critical boundary points*.[29,30] Each equation $\Psi_\alpha=0$ causes a line of the critical points of the first kind

$$x_{1*}' = x_1 - R(\mathbf{x}-\mathbf{x}_*')\sin\theta_*, \quad x_{2*}' = x_2. \tag{4}$$

The respective contribution of the isolated critical point of the first kind $\mathbf{x}_*' \in \Omega-\partial\Omega$ into integral (2) can be found by formula[29,30]

$$\mathbf{U}^*(\mathbf{x}) = -\frac{2\pi i}{k_*}\frac{1}{2\pi\mu}\mathbf{f}_*(\mathbf{x},\mathbf{x}_*')[\det\Psi''(\mathbf{x},\mathbf{x}_*')]^{-1/2} \times e^{ik_*\Psi(\mathbf{x},\mathbf{x}_*')} + O(k_*^{-2}). \tag{5}$$

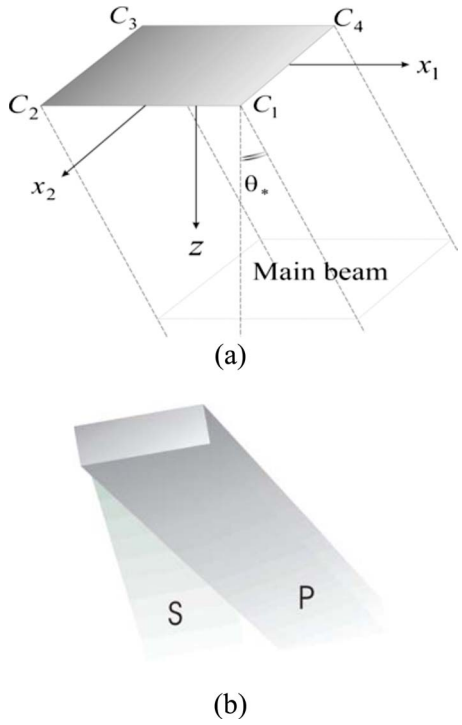


FIG. 3. (Color online) The main beams.

Neglecting the second order terms and substituting the magnitude, phase, and polarization vector into Eq. (5), we obtain

$$\mathbf{U}^*(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x}) = -\frac{ie^{ik_*(x_1\sin\theta_*+z\cos\theta_*)}}{\mu k_*\cos\theta_*} \times A_*(\theta_*)\mathbf{n}_*(\theta_*), \quad |x_1 - z\tan\theta_*| < a_1, \quad |x_2| < a_2, z > 0. \tag{6}$$

This is the half-space response to the normal stress $e^{ik_0x_1\sin\theta_0}$ given on its surface. However, when the critical point lies on the boundary ($\mathbf{x}_*' \in \partial\Omega$), the result halves,[29,30] and it equals zero outside the main beam [see Fig. 3(b)]. Since we deal with the asymptotic approach, this is not a contradiction, but an indication that $\partial\Omega$ is surrounded by a *boundary layer*, whose thickness is comparable with the neglected terms. We describe the boundary layers in Secs. IV–VI. Note that whether applying one dimensional stationary phase method to each one dimensional integral the result coincides with Eq. (6) from the 2D stationary phase method.[30]

Remark 1. The same approach for $\mathbf{U}^*(\mathbf{x})$ is applicable whatever the apodization $\mathbf{f}_*(\mathbf{x},\mathbf{x}_*')$, provided the latter contains no fast oscillating components (or k_*-independent). As in optics, the geometrical region where Eq. (6) applies is called *P*- or *S-main beam*, respectively [see Figs. 3(a) and 3(b)].

At *the critical points of the second kind* $\nabla\Psi(\mathbf{x},\mathbf{x}')$ may be nonzero, but $\nabla\Psi(\mathbf{x},\mathbf{x}')|_{\mathbf{x}'\in\partial\Omega} = \partial_{s'}\Psi(\mathbf{x},\mathbf{x}')|_{\mathbf{x}'\in\partial\Omega} = 0$, where s' is the curvilinear coordinate along $\partial\Omega$. Let us analyze the ribs $x_1' = \mp a_1$ and $x_2' = \mp a_2$ separately. The integral

$$\mathbf{I}_2 \equiv \int_{-a_2}^{a_2} \mathbf{f}_*(\mathbf{x},\mathbf{x}')e^{ik_*\Psi(\mathbf{x},\mathbf{x}')}dx_2' \tag{7}$$

can be evaluated using stationary phase method[29,30] to give

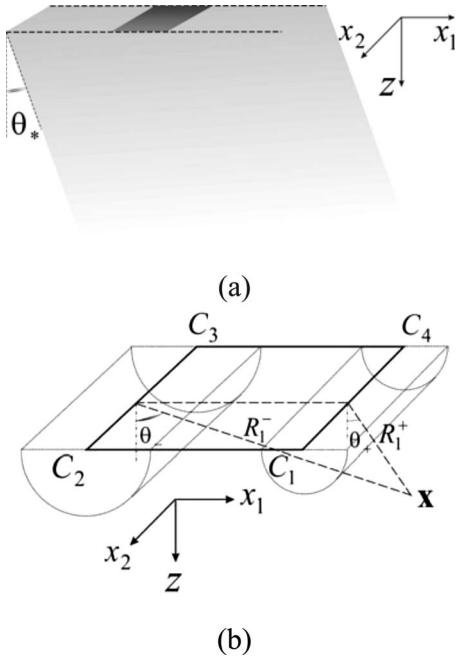


FIG. 4. (a) The side lobes containing the cylindrical edge waves and (b) diagram of waves.

$$\mathbf{I}_2 = \mathbf{I}_2(-a_2) + \mathbf{I}_2(\mathbf{x}'_{2*}) + \mathbf{I}_2(a_2) + O(k_*^{-\infty}), \tag{8}$$

where the contribution of $\mp a_2$ is analyzed below. The contribution $\mathbf{I}_2(\mathbf{x}'_{2*})$ of the "one dimensional" critical point of the first kind $\mathbf{x}'_{2*}=(x'_1,x'_{2*})$, $x'_2=x_2:|x_2|\le a_2$ of the leading asymptotic order has the form

$$\mathbf{I}_2(\mathbf{x}'_{2*}) = \sqrt{2\pi/k_*|\Psi_{22}(\mathbf{x},\mathbf{x}'_{2*})|}\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{2*}) \times e^{ik_*\Psi_*(\mathbf{x},\mathbf{x}'_{2*})+i\pi/4\ \mathrm{sgn}\ \Psi_{22}(\mathbf{x},\mathbf{x}'_{1*})} + O(k_*^{-1}),$$

$$\Psi_{22}(\mathbf{x},\mathbf{x}'_{2*}) = 1/R(\mathbf{x}-\mathbf{x}'_{2*}).$$

Using the partition of unity and integrating by parts, each rib $x'_1=\pm a_1$ contributes to

$$-\frac{1}{2\pi\mu}\int_{-a_1}^{a_1}\mathbf{I}_2(\mathbf{x}'_{2*})dx'_1 \to \frac{\mp 1}{ik_*2\pi\mu}\sqrt{\frac{2\pi}{k_*|\Psi_{22}(\mathbf{x},\mathbf{x}'_{2*})|}} \times\left\{\frac{\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{2*})}{\Psi_1}e^{ik_*(R_1+x'_1\sin\theta_*)+i\pi/4}+O(k_*^{-1/2})\right\},$$

$$R_1 = \sqrt{(x_1-x'_1)^2+z^2}$$

and results in the cylindrical edge wave as the leading order term of ray series

$$\mathbf{U}^*_{\mathrm{cyl}}(\mathbf{x}) = \pm\frac{e^{i3\pi/4}}{\mu\sqrt{2\pi k_*^3}}\frac{\sqrt{R_1^{\pm}}e^{ik_*(R_1^{\pm}\pm a_1\sin\theta_*)}}{R_1^{\pm}\sin\theta_*\pm a_1-x_1} \times A_*(\theta_\pm)\mathbf{n}_*(\theta_\pm), \quad R_1^{\pm}=\sqrt{(x_1\mp a_1)^2+z^2}. \tag{9}$$

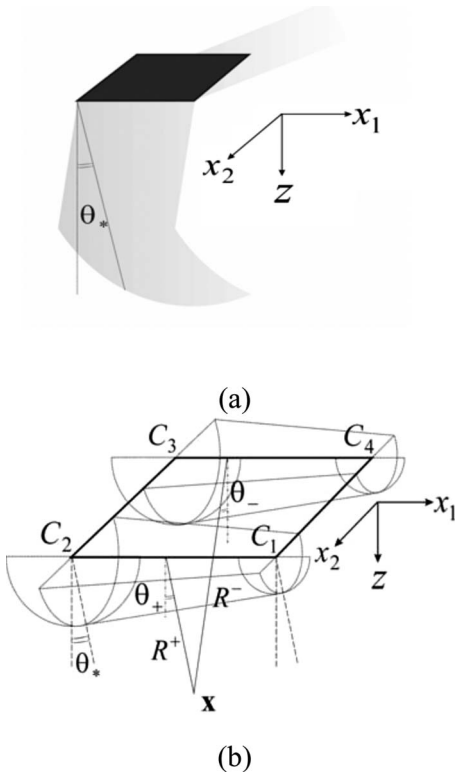


FIG. 5. (a) The side lobes containing the conical edge waves and (b) diagram of waves.

It is so called because its wave front (a surface of constant phase) is a *cylindrical surface*:

$$R_1^{\pm}\pm a_1\sin\theta_* \equiv \sqrt{(x_1\mp a_1)^2+z^2}\pm a_1\sin\theta_* = \mathrm{const}. \tag{10}$$

Thus, the half-layer $z>0$, $|x_2|\le a_2$ supports the *cylindrical edge waves* (9) from the ribs C_1C_4 and C_2C_3, each observation point inside the half-layer is reached by rays emanating from the point-source $x'_1=\pm a_1$, $x'_2=x_2$ [see Figs. 4(a) and 4(b)]. Above, $R_1^{\pm}$ is the distance between the source and observation point, and $\theta_\pm$ is the angle between $R_1^{\pm}$ and z-axis. Changing the order of integration in Eq. (2) we similarly obtain

$$\mathbf{I}_1 \equiv \int_{-a_1}^{a_1}\mathbf{f}_*(\mathbf{x},\mathbf{x}')e^{ik_*\Psi(\mathbf{x},\mathbf{x}')}dx'_1 = \mathbf{I}_1(-a_1)+\mathbf{I}_1(\mathbf{x}'_{1*}) + \mathbf{I}_1(a_1)+O(k_*^{-\infty}), \tag{11}$$

where the contribution $\mathbf{I}_1(\mathbf{x}'_{1*})$ of the critical point of the first kind $\mathbf{x}'_{1*}=(x'_{1*},x_2)$, $x'_{1*}=x_1-R(\mathbf{x}-\mathbf{x}'_{1*})\sin\theta_*$ is

$$\mathbf{I}_1(\mathbf{x}'_{1*}) = \sqrt{2\pi/k_*|\Psi_{11}(\mathbf{x},\mathbf{x}'_{1*})|}\{\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{1*}) \times e^{ik_*\Psi_*(\mathbf{x},\mathbf{x}'_{1*})+i\pi/4\ \mathrm{sgn}\ \Psi_{11}(\mathbf{x},\mathbf{x}'_{1*})}+O(k_*^{-1})\},$$

$$\Psi_{11}(\mathbf{x},\mathbf{x}'_{1*}) = R^{-1}(\mathbf{x}-\mathbf{x}'_{1*})\cos^2\theta_*,$$

$$-\frac{1}{2\pi\mu}\int_{-a_2}^{a_2}\mathbf{I}_1(\mathbf{x}'_{1*})dx'_2$$

$$\to \mp\frac{1}{ik_*2\pi\mu}\sqrt{\frac{2\pi}{k_*|\Psi_{11}(\mathbf{x},\mathbf{x}'_{2*})|}}$$

$$\times\left\{\frac{\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{2*})}{\Psi_2}e^{ik_*(R\cos^2\theta_*+x'_{1*}\sin\theta_*)+i\pi/4}+O(k_*^{-1/2})\right\}.$$

The leading term of the associated ray asymptotics is a *conical wave*

$$\mathbf{U}^*_{\mathrm{con}}(\mathbf{x})=\pm\frac{e^{i3\pi/4}}{\mu\sqrt{2\pi k_*^3\cos^3\theta_*}}\frac{\sqrt{R_2^{\pm}}}{\pm a_2-x_2}$$

$$\times e^{ik_*(R_2^{\pm}\cos\theta_*+x_1\sin\theta_*)}A_*(\theta_\pm)\mathbf{n}_*(\theta_\pm), \tag{12}$$

$$R^{\pm}=R|_{x'_2=\pm a_2}=\sqrt{(x_1-x'_{1*})^2+(x_2\mp a_2)^2+z^2}$$

$$=\sqrt{(R^{\pm}\sin\theta_*)^2+(R_2^{\pm})^2},\quad R_2^{\pm}=\sqrt{(x_2\mp a_2)^2+z^2}, \tag{13}$$

which is so called due to its wave front satisfying the equation of a *cone*

$$R_2^{\pm}\cos\theta_*+x_1\sin\theta_*=\mathrm{const}\to\sqrt{(x_2\mp a_2)^2+z^2}$$

$$+x_1\tan\theta_*=\mathrm{const}. \tag{14}$$

Thus, the region $z>0$ between two cones $x_1\mp a_1=R^{\pm}\sin\theta_*$ supports conical waves generated by C_1C_2 and C_3C_4. Each observation point inside these side lobes is reached by two rays from respective point-sources $x'_{1*}=x_1-R^{\pm}\sin\theta_*=x_1-R_2^{\pm}\tan\theta_*$, $x'_2=\pm a_2$ [see Figs. 5(a) and 5(b)]. Above, $R^{\pm}$ and $\theta_\pm$ denote the distance between the source and the observation point, and the angle between $R^{\pm}$ and z-axis, respectively. Clearly, when an observation point is close to the main beam lateral surface, i.e., $\pm a_1-x_1+R_1^{\pm}\sin\theta^*_{P,S}=0$ or $\pm a_2-x_2=0$, the expressions (9) and (12) are singular. Hence, the asymptotics (9) and (12) are valid everywhere except the vicinity of these surfaces where the boundary layer should be introduced.

Remark 2. When describing the edge waves, we integrated first by dx'_1 and then by dx'_2. It is easy to check that reversing the order gives the same result.

To obtain the corner waves let us apply the partition of unity to single out the contribution $\mathbf{I}_1(\pm a_1)$ to $x_1=\pm a_1$ in Eq. (11) and then do the same to $x_2=\pm a_2$ while integrating by dx'_2:

$$\mathbf{I}_1(\pm a_1)\cong\pm[ik_*\partial_{x'_1}\Psi(\mathbf{x},\mathbf{x}')]^{-1}\mathbf{f}_*(\mathbf{x},\mathbf{x}')e^{ik_*\Psi(\mathbf{x},\mathbf{x}')}|_{x'_1=\pm a_1}$$

$$+O(k_*^{-2}),$$

$$-\frac{1}{2\pi\mu}\int_{-a_2}^{a_2}\mathbf{I}_1(\pm a_1)dx'_2$$

$$\to\pm\pm\frac{1}{2\pi\mu}[k_*^2\partial_{x'_1}\Psi(\mathbf{x},\mathbf{x}')\partial_{x'_2}\Psi(\mathbf{x},\mathbf{x}')]^{-1}$$

$$\times\mathbf{f}_*(\mathbf{x},\mathbf{x}')e^{ik_*\Psi(\mathbf{x},\mathbf{x}')}|_{x'_1=\pm a_1,x'_2=\pm a_2}+O(k_*^{-2}).$$

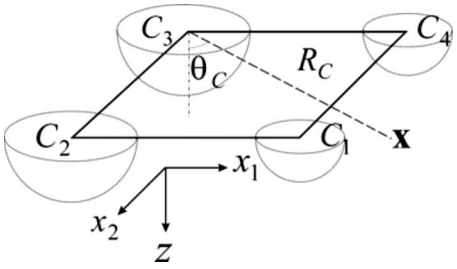


FIG. 6. The semi-spherical corner waves.

Hence, leading term of this contribution for C_{1-4}: $C(\pm a_1,\pm a_2)$ is

$$\mathbf{U}^*_c(\mathbf{x})=\frac{\mathrm{sgn}(x_{1C}x_{2C})}{2\pi\mu k_*^2}\frac{R_C}{(x_{2C}-x_2)(R_C\sin\theta_*+x_{1C}-x_1)}$$

$$\times e^{ik_*\Psi_C}A_*(\theta_C)\mathbf{n}_*(\theta_C),$$

$$R_C=\sqrt{r_C^2+z^2},\quad r_C=\sqrt{(x_1-x_{1C})^2+(x_2-x_{2C})^2}, \tag{15}$$

$$\Psi_C=R_C\pm x_{1C}\sin\theta_*,$$

whose wave front $\Psi_0=\mathrm{const}$ is a *semi-sphere* (see Fig. 6):

$$R_C=\mathrm{const}\mp a_1\sin\theta_*. \tag{16}$$

Again, when an observation point lies simultaneously on the boundary of main beam on the semi-sphere

$$\pm a_2-x_2=0,\quad R_C\sin\theta_*\pm a_1-x_1=0,$$

the formula (15) is singular and hence is no longer valid. Special boundary layer should be introduced in the vicinity of these singular points.

IV. BOUNDARY LAYERS

Above, we discussed the geometrical regions, which support a sum of finite number of waves, each described by ray series, i.e., the asymptotic series in inverse powers of the dimensionless wave number, with the coefficients depending on the observation point and the phase function (eikonal).[31] Such representations are not applicable inside the boundary layers[32,33] which are transition zones surrounding surfaces of irregularity along which rays of different nature meet. However, the fields inside boundary layers can be described by asymptotic series involving special functions (rather than just inverse powers), and the size and shape of a boundary layer are determined by the condition that the phase difference between coalescing rays is small. In our case this means that the critical point of the first kind from Ω tends to the lobe $x'_1=\pm a_1$ or $x'_2=\pm a_2$, i.e., its position is close to the critical point of the second kind. Constructing the respective boundary layer requires revising the convolution integral (2).

The penumbras associated with C_1C_2 and C_3C_4: Let us isolate the contribution $\mathbf{I}_1(\mathbf{x}'_{1*})$, $\mathbf{x}'_{1*}=(x'_{1*},x'_2)$, $x'_{1*}=x_1-R(\mathbf{x}-\mathbf{x}'_{1*})\sin\theta_*$ to the integral (11). Using the partition of unity, the contributions of vicinities $V_\varepsilon^-\equiv[-a_2,\varepsilon-a_2]$ and $V_\varepsilon^+\equiv[a_2-\varepsilon,a_2]$ are as follows:

$$\int_{-a_2}^{a_2} \mathbf{I}_1(\mathbf{x}'_{1*})dx'_2 \to e^{i\pi/4}\int_{V_\varepsilon^\pm} \sqrt{\frac{2\pi}{k_*|\Psi_{11}(\mathbf{x},\mathbf{x}'_{1*})|}} \times\{\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{1*})e^{ik_*\Psi(\mathbf{x},\mathbf{x}'_{1*})} + O(k_*^{-1})\}dx'_2.$$

Introducing the main beam phase Ψ_m, phase difference, and the new variable we obtain

$$\Psi_m = z\cos\theta_* + x_1\sin\theta_* = \Psi(\mathbf{x},\mathbf{x}'_{1*})|_{x_2=x'_2},\quad \Psi(\mathbf{x},\mathbf{x}'_{1*}) - \Psi_m = (R_2^\pm - z)\cos\theta_* \geq 0,$$

$$s(x'_2) \equiv \sqrt{k_*(R_2^\pm - z)\cos\theta_*}\,\mathrm{sgn}(x'_2 - x_2),$$

$$dx'_2 = \sqrt{\frac{R_2^\pm - z}{k_*\cos\theta_*}}\frac{2R_2^\pm}{|x'_2 - x_2|}ds \to \int_{V_\varepsilon^\pm}\cdots dx'_2 = \mp\int_0^{s_\varepsilon}\sqrt{\frac{(R_2^\pm - z)R_2^\pm}{k_*}}\frac{2}{|x'_2 - x_2|}A_*(\mathbf{x}-\mathbf{x}'_{1*})\mathbf{n}_*(\mathbf{x}-\mathbf{x}'_{1*})e^{is^2}ds.$$

The assumption $|x'_2 - x_2| \ll z$ and $x'_2 \approx \pm a_2$ yields

$$\sqrt{\frac{(R_2^\pm - z)R_2^\pm}{k_*}}\frac{2}{|x'_2 - x_2|} = \frac{2}{\sqrt{k_*}}\left(1+\frac{z}{R_2}\right)^{-1/2} = \frac{2}{\sqrt{k_*}}\left\{\frac{1}{\sqrt{2}} + \frac{1}{2\sqrt{2}}\left(\frac{x'_2 - x_2}{2z}\right)^2 + O\left(\frac{x'_2 - x_2}{2z}\right)^4\right\},$$

$$\mathbf{U}^*(\mathbf{x}) \cong -\frac{ie^{ik_*\Psi_0}}{\mu k_*\cos\theta_*}A_*(\mathbf{x}-\mathbf{x}_{\pm 2})\mathbf{n}_*(\mathbf{x}-\mathbf{x}_{\pm 2}) \times\frac{e^{-i\pi/4}}{\sqrt{\pi}}\int_{s_1}^{s_2} e^{is^2}ds,$$

$\mathbf{x}_{\pm 2} \equiv (x'_{1*}, \pm a_2)$. As $s_1 \to -\infty$, $k_* \to +\infty$ (for example, at $|x_2| \leq a_2$) we may conclude that

$$\frac{e^{-i\pi/4}}{\sqrt{\pi}}\int_{s_1}^{s_2} e^{is^2}ds \approx \frac{e^{-i\pi/4}}{\sqrt{\pi}}\int_{-\infty}^{s_2} e^{is^2}ds,$$

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

Then the leading term of penumbral asymptotics associated with $x'_2 \approx \pm a_2$ is expressed via *Fresnel integral* $F(\xi)$,

$$\mathbf{U}^*(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x})F(\sqrt{k_*(R_2^\pm - z)\cos\theta_*}\,\mathrm{sgn}(\pm a_2 - x_2)). \tag{17}$$

This boundary layer is so called by the analogy with a light-shadow zone in optics. Its shape is determined by the condition that the phase difference between the main beam and the edge wave equals π so that

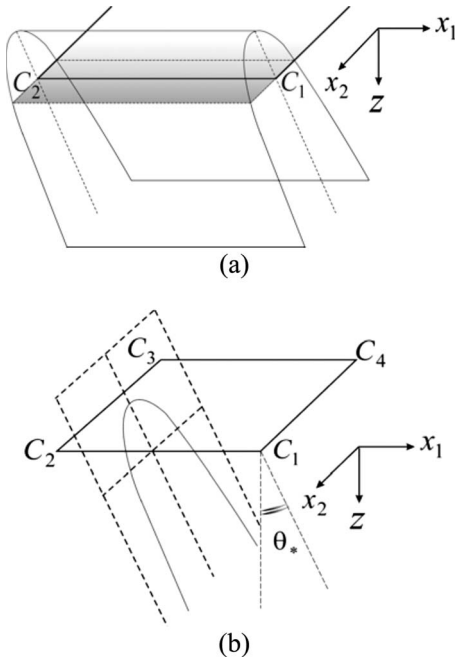


FIG. 7. (a) The penumbra associated with the edge C_1C_2 and (b) characteristic parabola.

$$\sqrt{k_*(R_2^\pm - z)\cos\theta_*}\,\mathrm{sgn}(\pm a_2 - x_2) = \sqrt{\pi} \to z = (x_2 \pm a_2)^2/2L_* - L_*/2, \tag{18}$$

where $L_* = \pi/k_*\cos\theta_*$ and $x'_{1*} = x_1 - R_2^\pm\tan\theta_*$. Hence, each of the above penumbras is a parabolic cylinder, declined by the angle θ_* with respect to z-axis [see Figs. 7(a) and 7(b)]. The larger the frequency is, the thinner these penumbras are.

The penumbras associated with C_2C_3 and C_4C_1: Similarly, consider the contribution $\mathbf{I}_2(\mathbf{x}'_{2*})$, $\mathbf{x}'_{2*} = (x'_1, x_2)$ to the integral (7) and analyze small vicinities $V_\varepsilon^- \equiv [-a_1, \varepsilon - a_1]$ and $V_\varepsilon^+ \equiv [a_1, a_1 - \varepsilon]$ in the integral

$$\int_{-a_1}^{a_1}\mathbf{I}_2(\mathbf{x}'_{2*})dx'_1 \to e^{i\pi/4}\int_{V_\varepsilon^\pm}\sqrt{\frac{2\pi}{k_*|\Psi_{22}(\mathbf{x},\mathbf{x}'_{2*})|}} \times\{\mathbf{f}_*(\mathbf{x},\mathbf{x}'_{2*})e^{ik_*\Psi(\mathbf{x},\mathbf{x}'_{2*})+i\pi/4} + O(k_*^{-1})\}dx'_1.$$

Introducing the new variables y and s we can write

$$y = x'_1 - x_1 + R_1\sin\theta_*,$$

$$R_1 = \frac{z^2 + y^2}{y\sin\theta_* + \sqrt{y^2 + z^2\cos^2\theta_*}},$$

$$R_{10} = R_1|_{y=0} = \frac{z}{\cos\theta_*},$$

$$\Psi(\mathbf{x},\mathbf{x}'_{2*}) - \Psi_m = \frac{y^2}{\sqrt{y^2 + z^2\cos^2\theta_*} + z\cos\theta_*} \sim \frac{y^2}{2z\cos\theta_*} \quad \text{as } |y| \ll z,$$

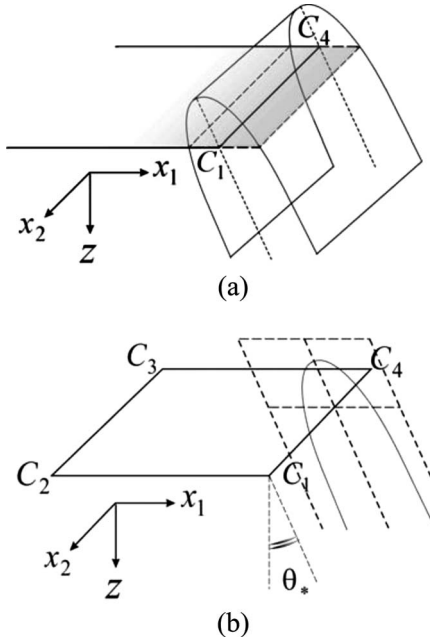


FIG. 8. (a) The penumbra associated with the edge C_4C_1 and (b) characteristic curve.

$$s = \sqrt{k_*(\Psi - \Psi_0)} = \sqrt{\frac{k_*}{\sqrt{y^2 + z^2\cos^2\theta_* + z\cos\theta_*}}}\, y\ \mathrm{sgn}\ y$$

$$\cong \sqrt{\frac{k_*}{2z\cos\theta_*}}\, y\ \mathrm{sgn}\ y,$$

$$ds = \frac{k_* y}{2sR_1}dx_1', \quad dx' = \frac{2sR_1}{k_* y}ds,$$

$$\mathbf{U}^*(\mathbf{x}) \approx -\frac{ie^{ik_*\Psi_0}}{\rho c_S^2 k_*\cos\theta_*}A_*(\mathbf{x}-\mathbf{x}_{\pm 1})\mathbf{n}_*(\mathbf{x}-\mathbf{x}_{\pm 1}) \times \frac{e^{-i\pi/4}}{\sqrt{\pi}}\int_{s_1}^{s_2} e^{is^2}ds, \quad \mathbf{x}_{\pm 1} \equiv (x_1 - R_1^{\pm}\tan\theta_* \mp a_1, x_2).$$

Simplifying, the leading penumbral term for $x_2 \in [-a_2, a_2]$ and the sides $\pm a_1$ equals

$$\mathbf{U}^*(\mathbf{x}) \approx \mathbf{U}_m^*(\mathbf{x})F(\sqrt{k_*\cos\theta_*(l_*^{\pm} - z)} \times \mathrm{sgn}(\pm a_1 \mp x_1 \pm R_1^{\pm}\sin\theta_*)), \quad (19)$$

$$l_*^{\pm} = \sqrt{(\pm a_1 \mp x_1 \pm R_1^{\pm}\sin\theta_*)^2 + z^2\cos^2\theta_*}/\cos\theta_*. \quad (20)$$

This penumbra differs from previous one and for $x_1' = x_1 - R_1\tan\theta_*$ its shape satisfies the condition that the phase difference between the main beam and the edge wave equals π,

$$\sqrt{k_*\cos\theta_*(l_*^{\pm} - z)}\,\mathrm{sgn}(\pm a_1 \mp x_1 \pm R_1^{\pm}\sin\theta_*) = \sqrt{\pi} \rightarrow l_*^{\pm} - z = \pi/k_*\cos\theta_*, \quad (21)$$

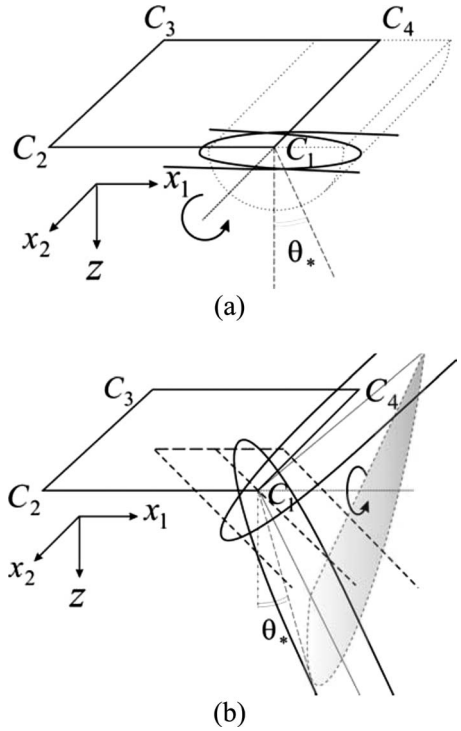


FIG. 9. (a) The edge-corner penumbra associated with C_1C_2 and corner C_1 and (b) the edge-corner penumbra for C_1C_4 and C_1.

$$z = \zeta^2/2L_* - L_*/2, \quad \zeta = (\mp a_1 - x_1 + R_1^{\pm}\sin\theta_*)/\cos\theta_*. \quad (22)$$

In the coordinates ζ, z, Eq. (22) is the equation of parabolic cylinder, but in coordinates x_1, z it is no longer so [see Figs. 8(a) and 8(b)]. When the angle is small (the case $\theta_* = 0$ corresponds to the normal probe) this shape is nearly the parabolic cylinder.

The edge-corner penumbra for C_1C_2 and corner C_1: As we have seen, for the cylindrical edge wave (9) the boundary layer should be introduced as well at $x_2 = \mp a_2$. Let us consider the interaction of boundary layers and suppose that the point-source at C_1C_2 tends to C_1. The contribution of the coalescing stationary point $x_1' = a_1$ to the integral (2) is treated similar to Eqs. (17) and (19). The leading term of ray asymptotics applicable in this *edge-corner penumbra* results in the expression

$$\mathbf{U}^*(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x})F(\sqrt{k_*\cos\theta_*(R_{C_1} - R_1^+)}\,\mathrm{sgn}(a_2 - x_2)), \quad (23)$$

$$R_{C_1} - R_1^+ = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + z^2} - \sqrt{(x_1 - a_1)^2 + z^2} \cong z[(x_2 - a_2)^2/2z^2 + O_4], \quad (24)$$

$$R_2^+ - z = \sqrt{(x_2 - a_2)^2 + z^2} - z \cong z[(x_2 - a_2)^2/2z^2 + O_4], \quad R_{C_1} - R_1^+ = R_2^+ - z + O_4, \quad (25)$$

where O_4 is the fourth order quantity with respect to $(x_2 - a_2)/z$. The shape of this boundary layer is determined by the condition that the phase difference between the edge and corner wave equals π,

$$k_* \cos\theta_*(R_{C_1} - R_1^+) = \pi. \tag{26}$$

The surface (26) shown in Fig. 9(a) is obtained by revolving the characteristic parabola in Fig. 7(b) about C_4C_1 in the lower half-space $z \geq 0$.

The edge-corner penumbra for C_4C_1 and corner C_1: By tending the point-source at C_4C_1 to C_1 the consideration of the coalescing stationary points yields

$$\mathbf{U}^*(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x}) F(\sqrt{k_* \cos\theta_*(R_*^+ - R_2^+)}\,\mathrm{sgn}(a_1 - x_1 + R_{C_1}\sin\theta_*)), \tag{27}$$

$$R_*^+ = \sqrt{\left(\frac{x_1 - a_1 - R_{C_1}\sin\theta_*}{\cos\theta_*}\right)^2 + (x_2 - a_2)^2 + z^2}, \tag{28}$$

$$R_*^+ - R_2^+ = z\left[\frac{1}{2}\left(\frac{x_1 - a_1 - R_{C_1}\sin\theta_*}{z\cos\theta_*}\right)^2 + O_4\right] = z\left[\frac{1}{2}\left(\frac{x_1 - a_1 - R_1^+\sin\theta_*}{z\cos\theta_*}\right)^2 + O_4\right], \tag{29}$$

$$l_*^+ - z = z\left[\frac{1}{2}\left(\frac{x_1 - a_1 - R_1^+\sin\theta_*}{z\cos\theta_*}\right)^2 + O_4\right], \quad R_{C_1} - R_1^+ = l_*^+ - z + O_4, \tag{30}$$

where O_4 is the fourth order quantity with respect to the coordinate difference (or difference from the rib to the main beam) from corner to depth ratio. The shape of the edge-corner penumbral layer is determined by the equation [see Fig. 9(b)]

$$k_* \cos\theta_*(R_*^+ - R_2^+) = \pi,$$

and this is a surface of revolution of the characteristic curve in Fig. 8(b) about C_4C_1 at $z \geq 0$. Other edge-corner boundary layers are obtained similar to Eqs. (23)–(30) by changing the coordinates of the respective corner and lobe.

V. THE CORNER BOUNDARY LAYERS

The shape of the penumbras associated with corners $C_1 - C_4$ is again followed from the requirement about the phase difference between the main beam and respective corner ray equals π. Introducing new variables rewrite the integral (2) as follows:

$$\begin{cases} \xi(x_1', x_2') = [x_1' - x_1 + R(\mathbf{x} - \mathbf{x}')\sin\theta_*]/\cos\theta_* \\ \eta(x_2') = x_2' - x_2 \end{cases}, \quad dx_1' dx_2' = [1 + O(\xi/z)^2 + O(\eta/z)^2]d\xi d\eta, \tag{31}$$

$$\mathbf{I} = -\int_{-a_2-x_2}^{a_2-x_2} d\eta \int_{\xi(-a_1)}^{\xi(a_1)} \frac{e^{ik_*\Psi_*}}{\mu k_*} \times \frac{k_*\cos\theta_* e^{ik_*\cos\theta_*(\xi^2+\eta^2)/2z}}{2\pi} \times \frac{A_*(\mathbf{x} - \mathbf{x}')\mathbf{n}_*(\mathbf{x} - \mathbf{x}')}{R(\mathbf{x} - \mathbf{x}')} d\xi. \tag{32}$$

Keeping in mind that an observation point is reached by a corner ray which lies closely to the main beam surface, let us simplify Eq. (32) and single out the main beam component

$$\frac{1}{R(\mathbf{x} - \mathbf{x}')} = \frac{1}{z\cos\theta_*} + O\left(\frac{\xi}{z^2}\right), \quad \frac{A_*(\mathbf{x} - \mathbf{x}')\mathbf{n}_*(\mathbf{x} - \mathbf{x}')}{A_*(\theta_*)\mathbf{n}_*(\theta_*)} = 1 + O\left(\frac{\xi}{z}\right),$$

$$\mathbf{I} \cong \frac{-iH}{\pi}\mathbf{U}_m^*(\mathbf{x}) \int_{-a_2-x_2}^{a_2-x_2} d\eta \int_{\xi(-a_1,x_2')}^{\xi(a_1,x_2')} e^{iH(\xi^2+\eta^2)} d\xi,$$

$$H \equiv \frac{k_*\cos\theta_*}{2z}.$$

For definiteness, consider the corner $C_1(a_1, a_2)$ whose contribution can be isolated by using the partition of unity[29,30] and extend the lower integration limits to $-\infty$,

$$\mathbf{I}_{C_1} \cong \frac{-iH}{\pi}\mathbf{U}_m^*(\mathbf{x}) \int_{-\infty}^{\eta_{C_1}} d\eta \int_{-\infty}^{\xi_{C_1}} e^{iH(\xi^2+\eta^2)} d\xi. \tag{33}$$

The area of integration in ξ, η is subdivided into the parts Ω_1' and Ω_1'' shown in Fig. 10(a). For region Ω_1' we pass to the polar coordinates r', ϕ' and obtain

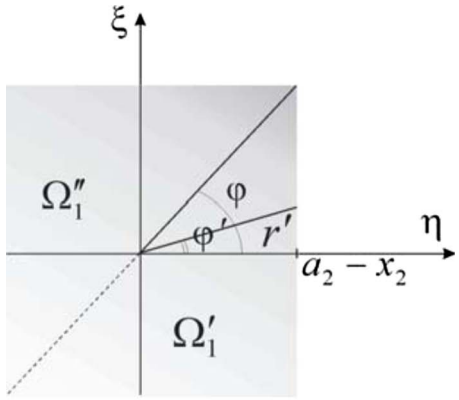
$$r'^2 = (\xi^2 + \eta^2)H, \quad \varphi' = \arg(\xi, \eta), \quad \varphi - \pi < \varphi' \leq \varphi,$$

$$r'(\varphi') = \begin{cases} (a_2 - x_2)\sqrt{H}/\cos\varphi', & -\pi/2 < \varphi' \leq \varphi \\ +\infty, & \varphi - \pi < \varphi' \leq -\pi/2, \end{cases}$$

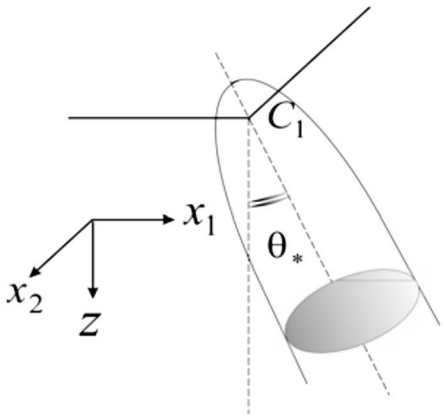
$$\mathbf{I}_{\Omega_1'} \cong -\frac{i\mathbf{U}_m^*(\mathbf{x})}{\pi} \int_{\varphi-\pi}^{\varphi} d\varphi' \int_0^{r'(\varphi')} e^{ir'^2} r' dr' = \mathbf{U}_m^*(\mathbf{x})\left\{\frac{1}{2} - \frac{1}{2\pi}\int_{-\pi/2}^{\varphi} e^{ir'^2} d\varphi'\right\}.$$

For the latter integral we introduce yet another set of variables and deduce

$$\nu'(\varphi') = r'(\varphi')\cos\varphi', \quad \lambda' = -\nu'\tan\varphi', \quad r'^2 = \lambda'^2 + \nu'^2, \quad d\varphi' = \nu' d\lambda'/(\lambda'^2 + \nu'^2),$$



(a)



(b)

FIG. 10. (Color online) (a) The integration region for corner C_1 in coordinates ξ, η and (b) penumbra associated with corner C_1.

$$\mathbf{I}_{\Omega_1'} \cong \mathbf{U}_m^*(\mathbf{x})\left\{\frac{1}{2} - G(-\nu'(\varphi)\tan\varphi, \nu'(\varphi))\right\}, \quad G(x,y) \equiv \frac{y}{2\pi}\int_x^{+\infty} e^{i(x'^2+y^2)^2}\frac{dx'}{x'^2+y^2}.$$

The function $G(x,y)$ is a *special function* called *generalized Fresnel integral*.[34] In the region Ω_1'': $\varphi < \varphi' \le \varphi + \pi$ we similarly obtain ($\varphi_2 = 0$),

$$r'(\varphi') = \begin{cases} \dfrac{\sin(\varphi-\varphi_2)\sqrt{(a_2-x_2)^2+\xi^2}}{\sin(\varphi'-\varphi_2)}, & \varphi < \varphi' \le \varphi_2 + \pi \\ +\infty, & \varphi_2 + \pi < \varphi' \le \varphi + \pi, \end{cases}$$

$$\mathbf{I}_{\Omega_1''} \cong -\frac{i\mathbf{U}_m^*(\mathbf{x})}{\pi}\int_\varphi^{\varphi+\pi} d\varphi' \int_0^{r'(\varphi')} e^{ir'^2} r'\,dr' = \mathbf{U}_m^*(\mathbf{x})\left\{\frac{1}{2} - \frac{1}{2\pi}\int_\varphi^{\varphi_2+\pi} e^{ir'^2} d\varphi'\right\},$$

$$\nu''(\varphi') = r'(\varphi')\sin(\varphi' - \varphi_2),$$

$$\lambda'' = -\nu''\cot(\varphi' - \varphi_2),$$

$$r'^2 = \lambda''^2 + \nu''^2, \quad d\varphi' = \nu'' d\lambda''/(\lambda''^2 + \nu''^2),$$

$$\mathbf{I}_{\Omega_1''} \cong \mathbf{U}_m^*(\mathbf{x})\left\{\frac{1}{2} - G(-\nu''(\varphi)\cot(\varphi-\varphi_2), \nu''(\varphi))\right\}.$$

Adding the above results we obtain

$$\mathbf{I}_{\Omega_1'} + \mathbf{I}_{\Omega_1''} \cong \mathbf{U}_m^*(\mathbf{x})\{1 - G(-\nu'(\varphi)\tan\varphi, \nu'(\varphi)) - G(-\nu''(\varphi)\cot(\varphi-\varphi_2), \nu''(\varphi))\}. \tag{34}$$

Note that the term 1 appears in Eq. (34) only if $\mathbf{x}$ is the inner point and, in fact, this is a Heaviside function. Thus, for C_1 the sum (33) yields

$$\mathbf{U}(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x})\{1_{C_1} - G(-\operatorname{sgn}(a_1 - x_1 + R_{C_1}\sin\theta_*) \times\sqrt{k_*\cos\theta_*(R_*^{\pm} - R_2^+)}, \operatorname{sgn}(a_2 - x_2) \times\sqrt{k_*\cos\theta_*(R_2^+ - z)}) - G(-\operatorname{sgn}(a_2 - x_2) \times\sqrt{k_*\cos\theta_*(R_*^{\pm} - l_*^{\pm})}, \operatorname{sgn}(a_1 - x_1 + R_{C_1}\sin\theta_*) \times\sqrt{k_*\cos\theta_*(l_*^{\pm} - z)})\}, \tag{35}$$

$$1_{C_1} = H(a_1 - x_1 + R_{C_1}\sin\theta_*)H(a_2 - x_2), \tag{36}$$

$$R_*^{\pm} = \sqrt{\left(\frac{\pm a_1 - x_1 + R_{C_1}^*\sin\theta_*}{\cos\theta_*}\right)^2 + (a_2 - x_2)^2 + z^2},$$

$$R_{C_{1,2}} = \sqrt{(a_1 \mp x_1)^2 + (a_2 - x_2)^2 + z^2}. \tag{37}$$

The shape of the boundary layer satisfies the equation

$$k_*\cos\theta_*(R^* - z) = \pi \tag{38}$$

and is a deformed paraboloid of revolution [Fig. 10(b)] whose axis is declined by angle θ_* from the vertical. For the normal transducer this is a straight paraboloid.

For the corner point C_2 we obtain similarly

$$\mathbf{U}(\mathbf{x}) \cong \mathbf{U}_m^*(\mathbf{x})\{1_{C_2} - G(-\operatorname{sgn}(a_1 + x_1 - R_{C_2}\sin\theta_*) \times\sqrt{k_*\cos\theta_*(R_*^{\pm} - R_2^+)}, \operatorname{sgn}(a_2 - x_2) \times\sqrt{k_*\cos\theta_*(R_2^+ - z)}) - G(-\operatorname{sgn}(a_2 - x_2) \times\sqrt{k_*\cos\theta_*(R_*^{\pm} - l_*^{\pm})}, \operatorname{sgn}(a_1 + x_1 - R_{C_1}\sin\theta_*)\sqrt{k_*\cos\theta_*(l_*^{\pm} - z)})\}, \tag{39}$$

$$1_{C_2} = H(a_2 + x_2 - R_{C_1}\sin\theta_*)H(a_2 - x_2). \tag{40}$$

The penumbras associated with C_3 and C_4 are symmetrical to those of corners C_1 and C_2, respectively.

VI. THE SUBCRITICAL AND SUPERCRITICAL INCIDENCE ANGLES

The above approximations represent the field as a sum of explicit *P*- and *S*-components: the plane waves (6), cylindrical and conical edge waves (9)–(12), and corner waves (15), each valid in the respective geometrical region. When these asymptotics become invalid we have to replace them by edge penumbral asymptotics (17)–(19) and edge-corner or

corner penumbral asymptotics (23), (27), (35), and (39). These asymptotics are nonuniform and the maximal asymptotic error is associated with the edge waves. They depend on the angle of incidence θ_0 (Fig. 1) which cannot be close to $\pi/2$ but θ_0 is often *supercritical* with respect to the inspection solid. If it is, all the results for *S -waves remain unchanged* but *no plane or edge conical P*-waves are present. The *edge cylindrical P*-waves (9) and *corner P*-waves (15) remain in place and become nonsingular. The *P-edge-corner penumbra* (23) does not change [see Fig. 9(a)], but *no other P*-penumbras exist.

VII. "EXACT" NUMERICAL SOLUTION

The exact numerical solution in the frequency domain, which can be obtained using double integral Fourier transform, results in the following integral:

$$\begin{bmatrix} U_\alpha \\ U_3 \end{bmatrix}(\mathbf{x}) = \frac{1}{4\pi^2\mu}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\begin{bmatrix} is_\alpha f \\ -pf_3 \end{bmatrix}\frac{T_3(\mathbf{s})}{R_0(\mathbf{s})} \times e^{i\mathbf{s}\mathbf{x}_0}ds_1ds_2, \quad \alpha = 1,2,$$

$$\begin{bmatrix} f \\ f_3 \end{bmatrix} = \begin{bmatrix} 2|\mathbf{s}|^2 - k_S^2 & -2pq \\ 2|\mathbf{s}|^2 - k_S^2 & -2|\mathbf{s}|^2 \end{bmatrix}\begin{bmatrix} e^{-pz} \\ e^{-qz} \end{bmatrix},$$

$$\mathbf{s} = (s_1, s_2), \quad \mathbf{x}_0 = (x_1, x_2),$$

$$R_0(\mathbf{s}) = (2|\mathbf{s}|^2 - k_S^2)^2 - 4pq|\mathbf{s}|^2, \quad p = \sqrt{|\mathbf{s}|^2 - k_P^2},$$

$$q = \sqrt{|\mathbf{s}|^2 - k_S^2}, \quad |\mathbf{s}|^2 = s_1^2 + s_2^2,$$

$$T_3(\mathbf{s}) = M(\omega)\int_{-a_2}^{a_2}\int_{-a_1}^{a_1} e^{i(k_0 x_1 \sin\theta_0 - \mathbf{s}\mathbf{x}_0)}dx_1dx_2 = 4M(\omega)\frac{\sin(k_0 \sin\theta_0 - s_1)a_1}{k_0 \sin\theta_0 - s_1}\frac{\sin s_2 a_2}{s_2}.$$

Similar representation based on the double Fourier transform has been obtained previously in Refs. 35 and 36. Due to the rapidly oscillating factors in the integrands, the numerical evaluation of the double integrals is extremely time consuming. In our code we have used special algorithms of the numerical algorithms group (NAG) subroutines.

VIII. SIMULATING THE PULSE TRAINS

The code based on the deduced asymptotics has been tested against the "exact" numerical solution. To model the propagation of pulses, we used the harmonic synthesis to simulate in each moment of time t. The displacement $\mathbf{U}(\mathbf{x},t)$ and particle speed are given by

$$\mathbf{U}(\mathbf{x},t) = \frac{1}{\pi}\mathrm{Re}\int_0^{+\infty} P(\omega)M(\omega)\mathbf{U}(\mathbf{x})e^{-i\omega t}d\omega, \quad \mathbf{V}(\mathbf{x},t) = \partial_t\mathbf{U}(\mathbf{x},t), \tag{41}$$

where the displacement spectrum $\mathbf{U}(\mathbf{x})$ has been obtained above, $P(\omega)$ is the Fourier transform of the load intensity, and $M(\omega)$ is a transfer function modeling the interaction of plastic wedge, fluid layer, and inspection solid. Below, the comparisons are performed for the most complicated case with the subcritical angle and wide-band pulse. The latter is chosen as 1 cycle of $Q\sin(2\pi ft)$ with $Q=500$ N, the carrier frequency $f=3$ MHz, and, consequently, the duration is 0.33 μs so that

$$P(\omega) = Q\int_0^{1/f}\sin(2\pi ft)e^{-i\omega t}dt.$$

It contains not only high but intermediate and low frequencies too. We compare the asymptotic code with the exact numerical code both combined with the fast Fourier transform to evaluate the integral (41). In the asymptotic code the bandwidth is taken to be 0.1–10 MHz, and in the exact one it is extended down to 0 MHz. The model parameters are as follows: in the plastic wedge wave speeds $c_P=2350$, $c_S=1150$ m/s, and mass density $\rho=1060$ kg/m^3; in the inspection solid $c_P=5961$, $c_S=3211$, and $\rho=7800$ (steel); and in the fluid layer of thickness h the wave speed $c=1653$ and mass density $\rho=1049$. The transducer dimensions are $a_1=10$ and $a_2=5$ mm. Typical illustration speed $v_3(t)$ normalized by its maximum value at the point $(0,0,H)$ is displayed in Fig. 11 where $\theta_0=15°$. In Fig. 11(a) $h=10^{-3}$ mm and solid line represents the asymptotics and dots indicate the pure numerical result. In our abbreviations the capital letters M, C, and E indicate the main beam, corner, and edge wave,

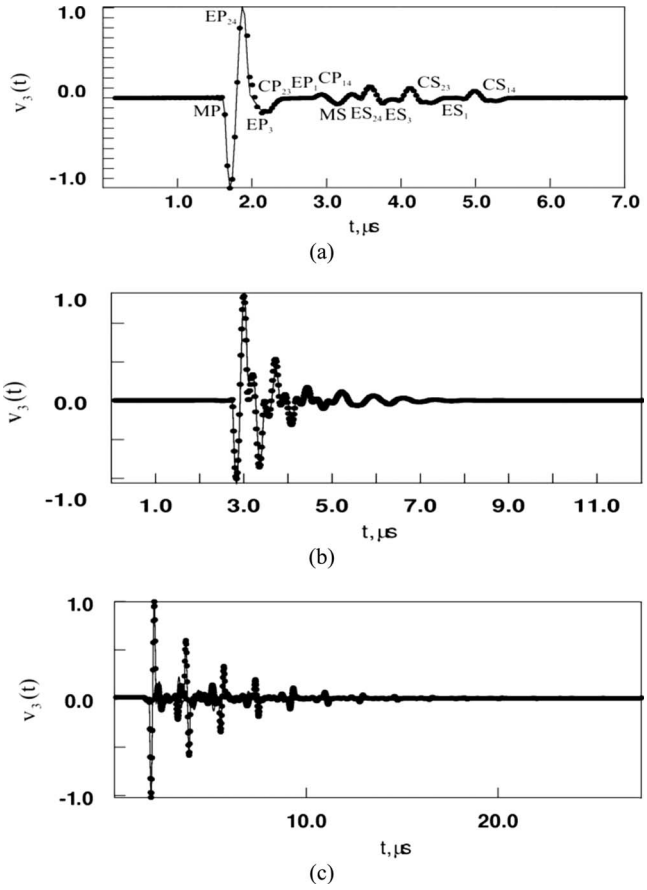


FIG. 11. The pulse trains in the polystyrene wedge/water/steel half-space. The wave arrivals are labeled by issuing corners and edges of contact spot, where the latter are (1) C_1C_4, (2) C_4C_3, (3) C_3C_2, and (4) C_2C_1.

respectively. The second letter P or S corresponds to the wave polarization. Each arrival is labeled according to its Huygens's source so that MP and MS arrivals originate at inner points of the contact spot, CP and CS at corner [the latter are numbered in Fig. 3(a)], and EP and ES at the spot edges. The depth $H=5$ mm is about five *S*-wavelengths. In Fig. 11(b) the fluid thickness is $h=0.3$, and in Fig. 11(c), $h=1.5$ mm. The time of the first arrival shifts with the fluid thickness. The response shown in Figs. 11(b) and 11(c) is more complicated than in Fig. 11(a) because of the multiple refractions, which result in overlapping the wave arrivals. However, the magnitudes of reflected pulses decrease rather fast. Therefore, if the fluid thickness is about 0.25–0.3 of the wavelength (a realistic assumption in many applications in nuclear industry), then multiple refractions do not complicate the result. The figure shows that the agreement between the exact and asymptotic codes is satisfactory, but using the same work station, the asymptotic code is $\approx 0.7\times 10^4$ faster.

IX. CONCLUSION

We have applied the two-tier asymptotic approach to evaluate the radiating near field of contact spot which models the angle beam and fluid coupled probe. The method relies on the high frequency asymptotics and the fast Fourier transform. It has been shown that in the geometrical regions underneath the contact spot the total radiated field contains several *P*- and *S*-components described in the approximate but explicit form. The edge waves are much smaller than waves in main beams. The corner waves are also small, but only these small signals are present inside certain zones. The penumbral zones have been described as well. The formulas elucidate the physics of the problem and give explicit dependence on its parameters. The code based on the deduced asymptotics has been validated against the exact numerical code. It proves to be $\sim 10^4$ times faster and provides a comparable accuracy even for the wide-band realistic pulses when most energy is concentrated at the higher end of the frequency band. Our approach represents an alternative to the Gaussian beam expansion[37] which is very computationally efficient and has been recently generalized to the case of rectangular normal probe.[38–40] However, it is based on the paraxial asymptotics and therefore is suitable only for modeling the main beam at a distance from the source. Our asymptotic approach may describe side lobes and penumbras as well and can be easily interfaced with other models of ultrasonic NDE. Note that we have not modeled the Rayleigh wave. This can be easily done by considering the contribution of the Rayleigh pole but is unnecessary when modeling the probes of consideration.

ACKNOWLEDGMENTS

This work has been carried out under EPSRC Grant Nos. GR/R17683/01 and GR/T22339/01 which are gratefully acknowledged. The authors are also thankful to Dr. Dmitri Gridin for sharing some of his codes and to Dr. Bob Chapman for useful discussions.

[1]A. J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials* (Springer, Berlin, 1990), pp. 145–150 and 196–202.
[2]T. P. Lerch, L. W. Schmerr, and A. Sedov, "The wavefield of an ultrasonic angle beam shear wave transducer: An elastodynamic approach," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. **16**. pp. 855–892.
[3]D. L. Lansing, "The displacements in an elastic half-space due to a moving concentrated normal load," NASA Technical Report No. R-238, National Aeronautics and Space Administration, Washington D.C., July 1966.
[4]D. C. Gakenheimer and J. Miklowitz, "Transient excitation of an elastic half space by a point load traveling on the surface," J. Appl. Mech. **36**, 505–515 (1969).
[5]M. C. M. Bakker, M. D. Verweij, B. J. Kooij, and H. A. Dieterman, "The traveling point load revisited," Wave Motion **29**, 119–135 (1999).
[6]K. J. Langenberg, U. Aulenbacher, G. Bollig, P. Fellinger, H. Morbitzer, G. Weinfurter, P. Zanger, and V. Schmitz, "Numerical modelling of ultrasonic scattering," in *Mathematical Modelling in Non-Destructive Testing*, edited by M. Blakemore and G. A. Georgiou (Clarendon, Oxford, 1988), pp. 132–138.
[7]U. Aulenbacher and K. J. Langenberg, "Transient radiation pattern of ultrasonic transducers on elastic half-spaces," in *New Procedures in Nondestructive Testing*, edited by P. Holler (Springer, Berlin, 1983), pp. 81–92.
[8]H. J. Stam, "The two-dimensional elastodynamic distributed surface load problem," Geophysics **55**, 1047–1056 (1990).
[9]L. W. Schmerr, Jr. and A. Sedov, "Elastodynamic model for compressional and shear wave transducers," J. Acoust. Soc. Am. **86**, 1988–1999 (1989).
[10]M. C. M. Bakker and M. D. Verweij, "The approximate far field and directivity of elastic wave transducers," J. Acoust. Soc. Am. **111**, 1177–1188 (2002).
[11]L. J. Fradkin, A. P. Kiselev, and E. Krylova, "The radiating near field asymptotics of a time-harmonic circular normal ultrasonic transducer in an elastic half-space," J. Acoust. Soc. Am. **104**, 1178–1187 (1998).
[12]D. Gridin and L. J. Fradkin, "The high-frequency asymptotic description of pulses radiated by a circular normal transducer into an elastic halfspace," J. Acoust. Soc. Am. **104**, 3190–3198 (1998).
[13]D. Gridin, "The radiating near field of a circular normal transducer of arbitrary apodization on an elastic half-space," J. Acoust. Soc. Am. **106**, 1237–1246 (1999).
[14]T. Xue, W. Lord, and S. Upda, "Numerical analysis of the radiated fields of ultrasonic transducers," J. Nondestruct. Eval. **14**, 137–146 (1995).
[15]A. Ilan and J. P. Weight, "The propagation of short pulses of ultrasound from a circular source coupled to an isotropic solid," J. Acoust. Soc. Am. **88**, 1142–1151 (1990).
[16]M. C. M. Bakker and M. D. Verweij, "Experimental validation for the wave field," J. Acoust. Soc. Am. **113**, 1850–1862 (2003).
[17]A. Lhémery, "A model for the transient ultrasonic field radiated by an arbitrary loading in a solid," J. Acoust. Soc. Am. **96**, 3776–3786 (1994).
[18]D. Gridin, "On the radiation of ultrasound into an isotropic elastic halfspace via wavefront expansions of the impulse response," J. Acoust. Soc. Am. **105**, 2565–2573 (1999).
[19]A. McNab, A. Cochran, and M. A. Campbell, "The calculation of acoustic fields in solids for transient normal surface force sources of arbitrary geometry and apodization," J. Acoust. Soc. Am. **87**, 1455–1465 (1990).
[20]F. G. Laturelle, "The stresses produced in an elastic half-space by a normal step loading over a circular area, analytical and numerical results," Wave Motion **12**, 107–127 (1990).
[21]L. F. Bresse and D. A. Hutchins, "Transient generation of elastic waves in solids by a disk-shaped normal force source," J. Acoust. Soc. Am. **86**, 810–817 (1989).
[22]H. Djelouah and J. C. Baboux, "Transient ultrasonic field radiated by a circular transducer in a solid medium," J. Acoust. Soc. Am. **92**, 2932–2941 (1992).
[23]J. C. Baboux and R. Kazys, "Analysis of the transient fields radiated in solids by circular and annular sources," J. Acoust. Soc. Am. **92**, 2942–2951 (1992).
[24]F. R. Norwood, "Exact transient response of an elastic half space loaded over a rectangular region of its surface," J. Appl. Mech. **36**, 516–522 (1969).
[25]F. Guan and M. Novak, "Transient response of an elastic homogeneous half-space to suddenly applied rectangular loading," J. Appl. Mech. **61**, 256–263 (1994).
[26]D. Gridin, "A fast method for simulating the propagation of pulses radiated by a rectangular normal transducer into an elastic half-space," J.

Acoust. Soc. Am. **104**, 3199–3211 (1998).
[27]D. D. Zakharov, "Approximate high order dynamic theory of a fluid layer in between two thick solids," J. Acoust. Soc. Am. **117**, 518–527 (2005).
[28]D. Gridin, "High-frequency asymptotic description of head waves and boundary layers surrounding the critical rays in an elastic half-space," J. Acoust. Soc. Am. **104**, 1188–1197 (1998).
[29]M. V. Fedoryuk, *Asymptotic Analysis* (Springer-Verlag, Berlin, 1992).
[30]R. Wong, *Asymptotic Approximations of Integrals* (Academic, New York, 1989).
[31]V. M. Babich and V. S. Buldurev, *Short-Wavelength Diffraction Theory: Asymptotic Methods* (Springer-Verlag, Heidelberg, 1991).
[32]R. N. Buchal and J. B. Keller, "Boundary layer problems in diffraction theory," Commun. Pure Appl. Math. **13**, 85–114 (1960).
[33]V. M. Babich and N. Ya. Kirpichnikova, *The Boundary-Layer Method in Diffraction Problems* (Springer-Verlag, Berlin, 1975).
[34]V. A. Borovikov, *Uniform Stationary Phase Method* (IEEE, Stevenage, Hertfordshire, UK, 1994).
[35]W. T. Thomson and T. Kobori, "Dynamic compliance of rectangular foundations on an elastic half-space," J. Appl. Mech. **30**, 579–584 (1963).
[36]D. V. Jones and M. Petyt, "Ground vibration in the vicinity of a rectangular load on a half-space," J. Sound Vib. **166**, 141–159 (1993).
[37]J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).
[38]H. J. Kim, L. W. Schmerr, and A. Sedov, "Generation of the basis sets for multi-Gaussian ultrasonic beam models—An overview," J. Acoust. Soc. Am. **119**, 1971–1978 (2006).
[39]D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gaussian beam expansion: Radiation fields of rectangular and elliptical transducers," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).
[40]K. Sha, J. Yang, and W.-S. Gan, "A complex virtual source approach for calculating the diffraction beam field generated by a rectangular planar source," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 890–897 (2003).

Using air-coupled sensors to determine the depth of a surface-breaking crack in concrete

Seong-Hoon Kee and Jinying Zhu[a)]
Department of Civil, Architectural, and Environmental Engineering, The University of Texas, Austin, Texas 78712-0273

(Received 20 May 2009; revised 16 October 2009; accepted 31 December 2009)

Previous studies showed that the surface wave transmission coefficient across a surface-breaking crack in concrete can be used to estimate the crack depth. However, inconsistencies in the surface wave transmission measurements limit the test accuracy and application of this technique. The inconsistencies come from near-field scattering by the crack tip and inconsistent sensor coupling conditions on rough concrete surfaces. This study first investigates the near-field size based on numerical analyses, and then suggests that reliable surface wave transmission should be measured in the far field. Based on the far-field measurement, the relationship between the surface wave transmission ratio and the normalized crack depth (crack depth/wavelength) is obtained. In the experimental study, the air-coupled sensing method is proposed as a solution to the sensor coupling problem. Owing to the non-contact feature, the air-coupled sensing method not only improves testing speed but also enables more consistent signal measurement. The experimental study using air-coupled sensors shows good agreement with the results of numerical simulation and analytic solution. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3298431]

PACS number(s): 43.35.Pt, 43.20.Hq, 43.20.El [ADP] Pages: 1279–1287

I. INTRODUCTION

The early stage of damage in reinforced concrete structures generally begins with various types of surface cracks. The appearance of cracks on the concrete members may not necessarily imply structural failure; however, it causes many serviceability and durability problems, e.g., stiffness degradation, corrosion of the reinforcement bars, and infiltration of moisture and/or deleterious ingredients. These effects cause further deterioration and lead to the early malfunction of concrete structures. From the viewpoint of civil infrastructure sustainability, it is important to monitor the extent of cracks in reinforced concrete members to evaluate the deterioration level and, if necessary, to make appropriate rehabilitation decision.

Using nondestructive evaluation (NDE) methods to estimate the depth of a surface-breaking crack has been investigated extensively in recent decades. Previous studies showed that the surface (Rayleigh) wave transmission coefficient Tr across a surface-breaking crack can be used to estimate the crack depth. Using ultrasonic transducers, Viktorov (1967) experimentally developed the relationship between the Rayleigh wave transmission coefficient and the normalized crack depth h/λ, i.e., the ratio of crack depth to the wavelength for a surface-breaking crack in a solid. Achenbach and his colleagues (Achenbach *et al.*, 1980; Mendelsohn *et al.*, 1980; Angel and Achenbach, 1984) analytically derived the Tr and h/λ relation based on diffraction and scattering of harmonic incident Rayleigh waves by a surface-breaking crack in far field. Masserey and Mazza (2005) extended the analytical solutions by prior researchers (Achenbach *et al.*, 1980; Mendelsohn *et al.*, 1980) to an arbitrary excitation function and verified that the established Tr and h/λ relation by Achenbach *et al.* (1980) is also valid for arbitrary incident waves. Yew *et al.* (1984) experimentally verified the Tr and h/λ relation for cracks in aluminum specimens, in which the surface waves were generated by dropping a steel ball on the surface of a specimen. For concrete, a heterogeneous but globally isotropic material, the surface wave transmission (SWT) method has been proved sensitive to depth variation in surface-breaking cracks, even for tightly closed and ill-defined cracks (Hevin *et al.*, 1998; Popovics *et al.*, 2000; Song *et al.*, 2003; Shin *et al.*, 2008). Hevin *et al.* (1998) obtained the transmission ratio of a surface wave in the frequency domain using boundary element analysis, and proposed the cut-off frequency ($h/\lambda=0.3$) method to estimate a crack depth. Popovics *et al.* (2000) and Song *et al.* (2003) obtained the Tr and h/λ relation based on experimental study in laboratory. They found that the relationship is not affected by the width of the crack opening and the shape of crack tips. Figure 1 shows the surface wave transmission coefficient versus normalized crack depth curves developed by these researchers. Overall, these curves show a similar trend that Tr decreases with increasing h/λ, but differences between the curves are also obvious. For example, at Tr=0.6, the crack depths h/λ given by these four curves can vary from 0.19 to 0.33. To successfully apply the SWT method, a reliable Tr versus h/λ curve should be determined first. In this study, the authors attempted to explain the discrepancies among these curves proposed by the previous researchers. Then based on the conclusions, we proposed a practical and reliable surface wave transmission method for evaluating the crack depth.

The discrepancies in previous research may be explained by the near-scattering effects of surface waves due to inter-

[a)]Author to whom correspondence should be addressed. Electronic mail: jyzhu@mail.utexas.edu

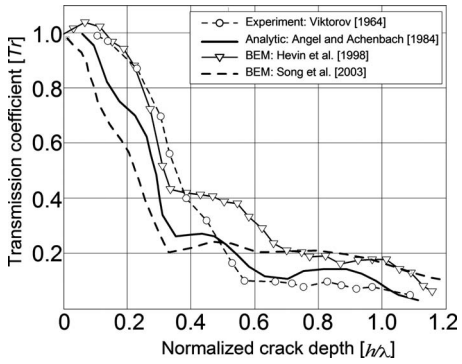


FIG. 1. Surface wave transmission coefficient versus normalized crack depth from previous literature.

action with the surface-breaking crack. Yew *et al.* (1984) observed that the signals measured in the near crack field were very complex due to the interference between surface waves and the bulk waves in the vicinity of a surface-breaking crack. Blackshire and Sathish (2002) investigated the near-field effect using a laser interferometer for a crack with $h/\lambda \gg 1$. They found that the signal amplitude at the crack location doubled the amplitude of the incident Rayleigh wave. The signal enhancement showed oscillatory variation in the vicinity of a surface-breaking crack and converged to a constant value when the measurement location is at a sufficiently far distance from the crack. Jian *et al.* (2007) studied the near-field scattering of the Rayleigh wave using numerical simulations [finite element method (FEM)]. In that investigation, they explained that the signal enhancement was mainly due to the interference effect of bulk waves (i.e., mode converted *P*- and *S*-waves in front of the crack, and bulk waves generated from the crack tip) and direct contribution of incident surface waves. Therefore, the surface wave transmission measurements will be affected by the near-field effect when the sensors are too close to the crack. In practice, however, the sensors cannot be placed too far from the crack (in true far field) due to dimension limit of the test specimen. This study investigated near-field effect variation with the sensor-to-crack spacing and proposed an approximate far-field sensor arrangement guideline.

In experiments, Achenbach *et al.* (1992) proposed a self-calibrating technique to reduce the experimental variability in surface wave measurement. The surface wave transmission ratios were measured from two opposite directions and averaged (the specific procedures are shown in the experimental section of this paper). This setup can eliminate effects of source variation and the coupling difference between two sensors, based on the assumption that coupling of the sensors during tests remains consistent (Popovics *et al.*, 2000). However, temporary sensor mounting methods, such as wax, adhesive, and magnetic mounts, will induce a low resonant frequency (usually below 10 kHz) which depends on the mounting stiffness of the contact sensors (e.g., accelerometers). The mounting resonant response affects the repeatability of tests. Poor signal consistency has been noticed in low frequency range by the authors and other researchers (Hevin *et al.*, 1998). The errors caused by this mounting resonance cannot be eliminated by the self-calibrating technique. On the other hand, permanent mounting cannot meet the requirement of rapid NDT tests and is difficult to apply to concrete. This study proposed a solution to this problem by using non-contact air-coupled sensors to improve test accuracy and efficiency.

Non-contact ultrasonic sensing techniques for concrete include air-coupled sensing and laser ultrasonic techniques. The laser technique has been shown effective to characterize Rayleigh wave attenuation in cement-based materials (Jacobs and Owino, 2000). However, the application of laser ultrasonic technique is still very limited in field testing due to expensive equipment and low reflection from rough surface of concrete. Air-coupled sensors have been successfully used to measure leaky surface waves in concrete by Zhu and Popovics (2005). Compared to contact sensors, the air-coupled sensors have the following benefits: (i) the non-contact sensing technique eliminates sensor coupling problems and thus gives more consistent measurement results; (ii) the non-contact feature enables rapid scanning of large civil engineering structures. To improve signal quality and test efficiency, this study combines the air-coupled sensing with the self-calibrating setup in surface wave transmission measurements.

This study first uses numerical simulations to investigate the parameters affecting surface wave transmission measurements, which includes the depth of crack h, impact duration of the source, and locations of sensors. An approximate far-field size is proposed. The Tr and h/λ relationship, based on measurements in the far field, is obtained. Then a simplified algorithm to calculate surface wave transmission coefficient is proposed in Sec. II E. In the experimental study, surface wave transmissions were measured from a series of notch-typed cracks in a concrete slab using air-coupled sensors. The experimental results validate the numerical results and show the promising application of air-coupled sensors.

II. NUMERICAL SIMULATION

A. FEM

The FEM was used to simulate the propagation behavior and near-scattering of surface waves caused by a surface-breaking crack in concrete. To save computational cost, a two-dimensional (2D) model was developed using rectangular quadratic axisymmetric elements (CAX4) in the commercial program (ABAQUS Inc., 2007). The validity of the model has been verified by Kim and Kwak (2008) to investigate properties of concrete using surface waves generated by a point impact source.

The model mesh size l_e and the time increment Δt for numerical integration procedure were carefully determined based on the wave velocity and frequency content (Ihlenburg, 1998). The mesh size was designed as 5 mm so that at least ten elements can participate to express the minimum wavelength λ_{min} in this study (Alleyne and Cawley, 1991). In this study, the λ_{min} for shear waves in concrete at $f_{max}=50$ kHz is about 50 mm. The maximum time increment Δt for integration was determined as 1 μs, which is smaller than the characteristic time τ ($\tau=l_e/V_p$, where V_p is *P* wave velocity of concrete). Thus, a disturbance cannot propagate

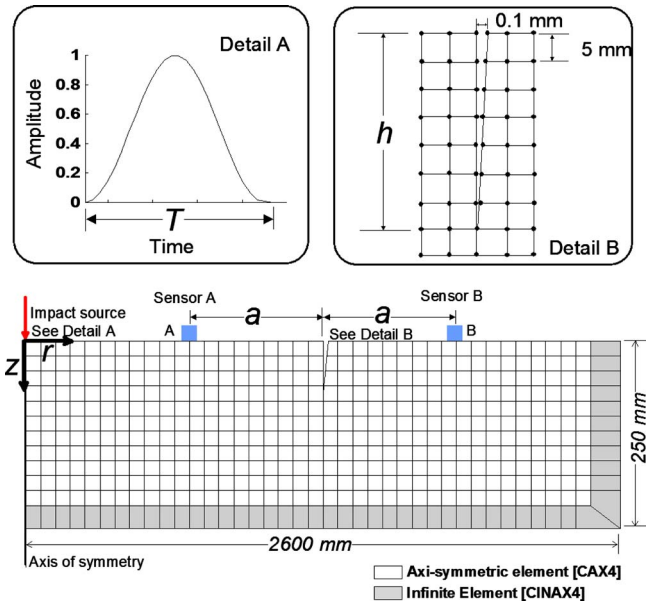


FIG. 2. (Color online) Finite element model for elastodynamic transient analysis in a solid half-space.

through a grid size during one time step. The finite element model is shown in Fig. 2. Sixteen models were analyzed for crack depth h ranging from 0 to 150 mm in a 10 mm step. To reduce wave reflections from boundaries, absorption boundary (CINAX4) elements were applied. The material properties are Young's modulus of 33 630 MPa, Poisson's ratio of 0.22, and mass density of 2400 kg/m^3. The corresponding velocities of P-, S-, and surface waves are 4050, 2420, and 2215 m/s, respectively. These material properties were chosen to simulate normal concrete.

A transient impact point source was applied on free-surface of the half-space at the axis of symmetry (i.e., $r=0$). The force function of the source is

$$f(t) = \begin{cases} \sin^2(\pi t/T), & 0 \le t \le T \\ 0, & t > T, \end{cases} \quad (1)$$

where T is the duration of impact. Note that the quadratic force function was selected to reduce Gibbs effect on FEM analysis (Kim and Kwak, 2008). The impact source is 1300 mm from the crack opening on the surface. To study the near-field effect caused by the crack tip scattering, various sensor locations on the surface of the concrete were investigated.

B. Near-field scattering of surface waves due to a surface-breaking crack

The size of near-scattering field around a crack depends on the crack depth and wavelength of the incident surface waves. Six numerical models, including two crack depths h =10 and 70 mm and four impact duration time T=40, 80 and 120 μs, were analyzed. These models correspond to six different h/λ of 0.12, 0.81, 0.06, 0.41, 0.04, and 0.27. Figure 3 shows the B-scan images of vertical velocity responses on the solid surface from the six models.

When h is much smaller than λ, e.g., h/λ=0.12 in Fig. 3(a), the dominant wave components are the reflected surface

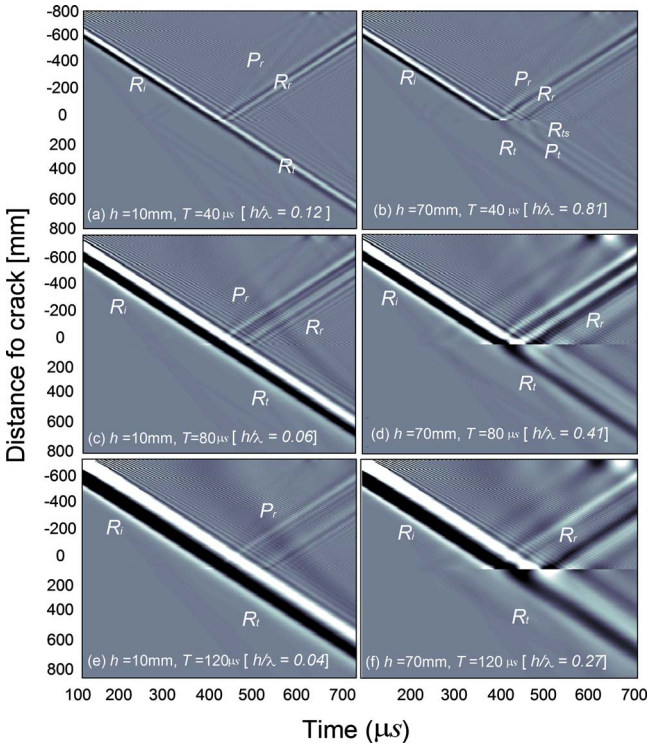


FIG. 3. (Color online) B-scan images of wave propagating on solid models with various h/λ: (a) 0.12, (b) 0.81, (c) 0.06, (d) 0.41, (d) 0.04, and (e) 0.27. (a)–(f) represent six models with two crack depths h=10 mm [(a), (c), and (e)] and 70 mm [(b), (d), and (f)], and three source duration times T =40 μs [(a), and (b)], 80 μs [(c), and (d)], and 120 μs [(e) and (f)].

waves (R_r) and the transmitted surface waves (R_t) after the incident waves (R_i) interact with the crack. The low frequency components of the surface waves can transmit through the crack, while most of high frequency components are reflected back by the crack. On the other hand, when $h/\lambda \sim 1.0$ (deep crack), most portions of the incident waves are reflected by the crack, as shown in Fig. 3(b).

In Figs. 3(a) and 3(b), the mode converted P-waves with a velocity of approximately 4000 m/s (P_r) are observed along with R_r. However, when the impact duration T increases to 80 and 120 μs, as shown in Figs. 3(c)–3(f), the mode converted bulk waves are not distinguishable from surface waves. Interferences between the incident and reflected surface waves are observed in the near region of the crack, as shown in Figs. 3(a) and 3(b). In addition to the reflected and transmitted surface waves, mode converted bulk waves and surface-skimming waves are generated by crack tip scattering.

C. Signal enhancement in the near field

Interaction between the mode converted waves (bulk waves, and multi-reflected waves) and the incident surface waves induces signal enhancement in the forward and backward scattering fields. This phenomenon has been noticed by Jian *et al.* (2007); however, the affected range (near-field size) was not studied in detail. In this paper, the authors present the signal enhancement around a crack and affecting parameters. For this purpose, the amplification coefficient (APC) is defined as the peak amplitude ratio between the

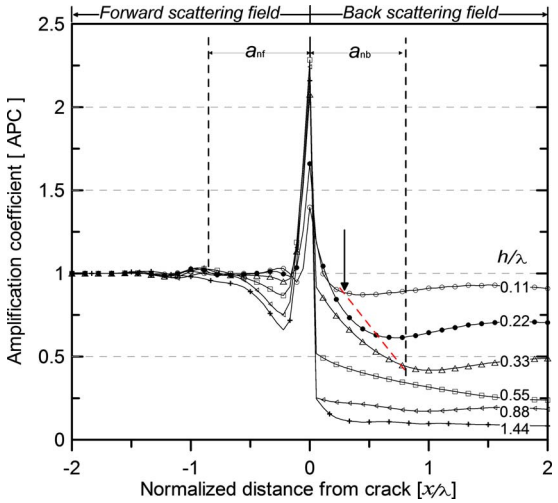


FIG. 4. (Color online) APC varies with the normalized distance from crack (x/λ) for different h/λ.

vertical displacements obtained on the cracked model (U_{zi}) and that on the crack free model (U_{z0}) as follows:

$$\mathrm{APC}(x/\lambda) = \frac{U_{zi}(x/\lambda)}{U_{z0}(x/\lambda)}. \tag{2}$$

APC was calculated for models of various impact durations T of 40–140 μs and crack depths h of 0~150 mm. Analysis based on the wide range of T and h shows that the APC depends on the distance from the crack x, the crack depth h, and the wavelength λ. Figure 4 shows amplification coefficient curves versus the normalized distance x/λ from the crack opening for various normalized crack depths (h/λ). In the backward scattering field, APC curves show very small oscillation (1%) when x is larger than 0.2λ. In the forward scattering field, APC curves decrease sharply from the upper peak at the location of the crack and then gradually reach a constant value with increasing x/λ. This constant value is the surface wave transmission coefficient in the far field.

Based on the results shown in Fig. 4, in the range of $h/\lambda = 0 \sim 1/3$, the approximate near-field size a_n measured from the crack opening can be expressed as

$$a_n/\lambda = 1.8h/\lambda + 0.1 \quad \text{for} \quad 0 \le h/\lambda \le 1/3. \tag{3}$$

Cheng and Achenbach (1996) suggested that $x/\lambda > 5$ can be approximately regarded as in the far field. However, this far-field requirement is difficult to satisfy in experiments due to the specimen size limit. Equation (3) is a less strict criterion, which provides a guideline to avoid significant signal enhancement. Surface wave transmission based on this criterion is slightly different from the curve built on far-field analysis (Angel and Achenbach, 1984), which will be shown in the following Sec. II D.

D. Calculation of transmission coefficient

The surface wave transmission function across a surface-breaking crack is defined as

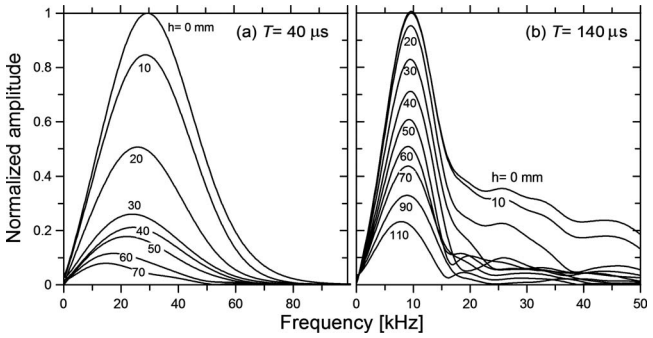


FIG. 5. Frequency spectra of windowed signals for various crack depths. The impact source duration times are (a) 40 μs and (b) 140 μs.

$$\mathrm{Tr}(f,h) = V_b(f,h)/V_a(f,h), \tag{4}$$

where Tr is the transmission coefficient in function of frequency and crack depth, and V_a and V_b are the Fourier transforms of the time domain signals measured at locations of A and B, respectively (Fig. 2). Sensors A and B are located in the far field based on the criterion in Eq. (3). To eliminate geometric attenuation effects, the transmission coefficient $\mathrm{Tr}(f,h)$ is normalized by the value obtained from a crack free model $\mathrm{Tr}(f,0)$.

Figure 5 shows the signal spectra $V_b(f,h)$ corresponding to impact source durations T of 40 and 140 μs, respectively, for crack depths h varying from 0 to 150 mm. A window function (a rectangular window) of three times of the impact duration $(3T)$ is applied to time domain signals to extract the transmitted surface wave signals. The figure shows that the center frequency f_c, the maximum amplitude, and spectral energy (area under the spectral curve) of the *transmitted surface waves* decrease with increasing crack depth h.

Figure 6 shows the normalized transmission coefficient Tr_n versus h/λ relationship when sensors are located at 650 mm $(\sim 5\lambda)$ from the crack opening. The sensors are in far-field region, according to Cheng and Achenbach (1996). The data in Fig. 6 were obtained from models with $T=60$ μs (center frequency of 17 kHz) and h from 0 to 100 mm. To investigate the effect of Poisson's ratio υ, transmission curves for υ of 0.2 and 0.3 were calculated. The results are shown as lines with open squares and open circles in Fig. 6,

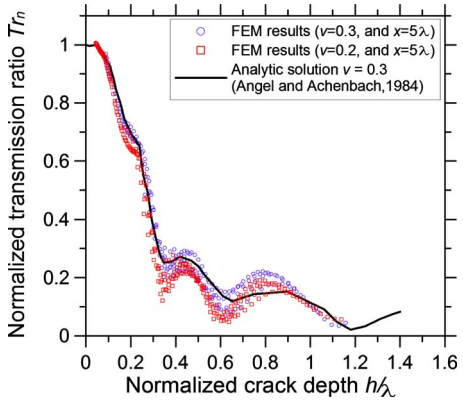


FIG. 6. (Color online) Normalized transmission coefficient versus normalized crack depth relation based on FE models and the analytic solution in far-field regions.

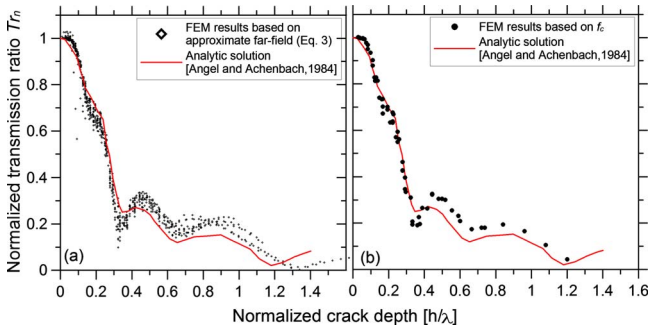


FIG. 7. (Color online) Normalized transmission coefficient versus normalized crack depth relation based on FE models in the approximate far-field region. The far-field analytic solution is also shown.

respectively. For comparison purposes, the analytical result given by Angel and Achenbach (1984) for v of 0.3 is also shown as solid line in Fig. 6. The transmission curve for $v=0.3$ shows good agreement with the analytical curve, which validate the numerical analysis models used in this study. Figure 6 also shows that Poisson's ratio has slight effect on the transmission function, but it is insignificant compared to other critical factors (h, x, and λ).

Figure 7(a) shows the Tr_n and h/λ relationship based on FEM models with T from 40 to 140 μs and h from 0 to 150 mm. The sensors are located in the approximate far-field region given by Eq. (3). For example, for $T=60$ μs and $h=50$ mm, the sensor location x was 110 mm ($a_n \sim 103.5$ mm with $\lambda \sim 135$ mm for $T=60$ μs). Overall, the data in Fig. 7(a) indicate that the FEM results agree with the analytic curve well, especially in the range of $h/\lambda=0 \sim 1/3$. Scattering of FEM data may come from two aspects: different Poisson's ratio (0.22 versus 0.3) and near-field effects. The criterion given by Eq. (3) is just an approximation to the true far field. Nevertheless, the data scattering has been significantly decreased when it is compared with previous research without considering near-field effects (Hevin *et al.*, 1998). The near-field effect may be a possible reason why the transmission curves obtained by other researchers (Song *et al.*, 2003) are different from the analytical results.

E. Simplified algorithm

The crack depth can be estimated by inversion process of the measured surface wave transmission function based on the established transmission curve. Reliable surface wave transmission measurements are the key to accurate evaluation of the crack depth. Figure 5 shows that a signal has the highest energy at the center frequency. Therefore, measurements around the signal center frequency may provide the most reliable results. Based on this assumption, we proposed a simplified algorithm to calculate the surface wave transmission. Instead of calculating wave transmission for all frequencies, we calculated Tr at the center frequency only. Then Eq. (4) is modified to

$$\mathrm{Tr}_n(f_c,h) = \mathrm{Tr}(f_c,h)/\mathrm{Tr}(f_c,0), \tag{5}$$

where f_c is center frequency of the input signal $V_a(f,0)$ (Fig. 5).

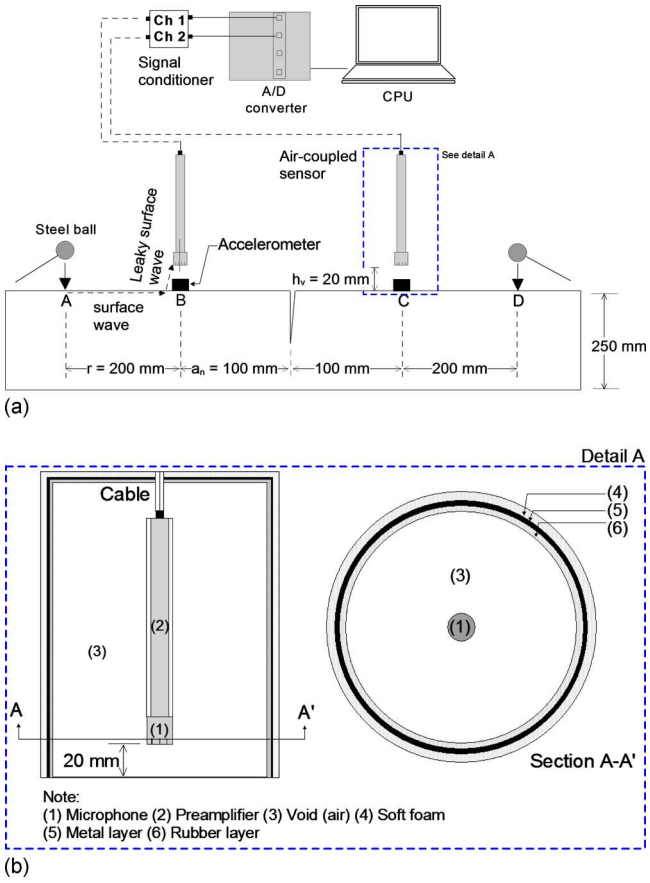


FIG. 8. (Color online) (a) Test setup based on self-compensation technique. (b) Detail of air-coupled sensor with sound insulation device.

With the measured surface wave transmission ratio at the center frequency, h/λ can be directly found from the established $\mathrm{Tr}_n \sim h/\lambda$ curve. If the surface wave velocity is known, the wavelength λ can be calculated, and then the crack depth can be determined. Figure 7(b) shows the $\mathrm{Tr}_n \sim h/\lambda$ curve based on measurements at center frequencies f_c. The impact source duration varies from 40 to 140 μs, and the crack depth h varies from 0 to 150 mm at 10 mm intervals. Good agreement is also observed between the center frequency measurement and the analytic solution, especially in the sensitive region where $h/\lambda \leq 1/3$.

III. EXPERIMENTAL VERIFICATION

A. Test setup

The surface wave transmission coefficient was measured for a notched concrete specimen using the air-coupled self-calibrating scheme. A schematic view of the test setup is shown in Fig. 8(a). Two air-coupled sensors were used to measure leaky surface waves propagating along the concrete surface. The air-coupled sensors (PCB model 377B01) have the following properties: nominal diameter of 6 mm, ± 2 dB flat frequency response over 4 Hz to 80 kHz, and resonance frequency around 83 kHz. The horizontal spacing between the sensor and the impact source is designed as 200 mm to minimize the effects of direct acoustic waves. The sensor-to-crack spacing is 100 mm according to Eq. (3) to reduce near-field effects. According to Zhu and Popovics (2005), the

air-coupled sensors should be close to the test surface to reduce ambient noise and direct acoustic wave effects. In this study, the vertical distance between the sensors and the concrete surface h_v is 20 mm. Unlike a contact sensor, which measures waves right at the sensor location, an air-coupled sensor actually measures the surface waves emitting from the concrete surface at a distance of $h_v \tan\theta$, where θ is the leaky angle determined by Snell's law. This distance is defined as shadow zone. When this value is large, sensor locations should be corrected for shadow zone. In this study, the shadow zone size is about 3.0 mm, which is very small compared to the sensor to crack spacing at 100 mm.

To further reduce effects of direct acoustic waves and ambient noise, the air-coupled sensor was shielded by a sound insulation device, as shown in Fig. 8(b) (Zhu and Popovics, 2005).

A 250 mm thick concrete slab was cast in the laboratory. The thickness of the slab is larger than the wavelength of surface waves used in this study, so that the slab can be approximately regarded as a half-space, as assumed in the analytical and numerical models. Notch-type cracks with depth varying from 10 to 100 mm were generated in the specimen by inserting a 0.5 mm thick metal sheet before casting concrete. The metal sheet was removed from the concrete 12 h later. The width of the notches was measured by a crack width gauge, and it is about 0.5 mm. So the width-to-depth ratio of the notches used in this study is smaller than 0.1. According to Masserey and Mazza (2005), narrow slots with width-to-depth ratio smaller than 0.1 can be regarded as a crack, so that results from experimental study may be directly comparable with the theoretic results. The concrete was made from normal Portland cement type I/II, river sand, and gravel with a maximum size of 10 mm. The density of concrete is 2350 $\mathrm{kg/m^3}$, which was measured from five concrete cylinders (diameter of 10 cm and height of 20 cm), cast at the same time with the concrete specimen. The phase velocity of surface waves was calculated from a dispersion curve using the multichannel analysis of surface waves method (Ryden and Park, 2004). The velocity converged to 2200 m/s when the frequency is greater than 10 kHz. The velocities of *P*- and *S*-waves measured by low frequency ultrasonic transducers (50 kHz) in a through transmission mode were 4100 and 2460 m/s, respectively.

B. Data acquisition

The self-calibrating technique was used to measure surface wave transmission across the notch in concrete. The test setup is shown in Fig. 8(a). First, surface waves generated by an impact load at location A were recorded by sensors at locations B and C. The signals are denoted as V_{AB} and V_{AC}. Consequently, the wave transmission between locations B and C can be calculated from these two signals and denoted as T_{BC}. To eliminate the unsymmetrical effect caused by sensor coupling, an impact load is then applied at location D, and the signals recorded by the sensors at locations B and C are denoted as V_{DB} and V_{DC}. The transmission ratio between locations C and B is defined as T_{CB}. The average surface wave transmission function is (Popovics *et al.*, 2000)

$$|T_{BC}(f)| = \left| \sqrt{\frac{V_{AC}V_{DB}}{V_{AB}V_{DC}}} \right|. \tag{6}$$

To improve signal consistency, five repeated signal data sets were collected at the same test location. These five transmission functions were then arithmetically averaged in frequency domain. The signal coherence function shown in Eq. (7) was used to evaluate the consistency of obtained signals

$$SC_{BC}(f) = \frac{\left|\sum G_{BC}(f)\right|^2}{\sum G_{BB}(f) \times \sum G_{CC}(f)}, \tag{7}$$

where $G_{BC}(f)$, $G_{BB}(f)$, and $G_{CC}(f)$ are the cross spectrum and autospectrum functions between V_{AB} and V_{AC}, respectively. Similarly, $SC_{CB}(f)$ can also be calculated from signals V_{DB} and V_{DC}. The averaged $SC(f)$ is defined as

$$SC(f) = SC_{BC}(f)SC_{CB}(f), \tag{8}$$

which ranges from 0 to 1.0. A value close to 1.0 indicates good signal quality and repeatability. Therefore, the signal coherence function can be used to select the acceptable frequency range of a transmission ratio curve.

The measured surface wave transmission ratio Tr was further normalized by Tr_0, which is the transmission ratio obtained from a crack free region. This procedure will eliminate the geometric effect caused by a point source. All analyses were performed in the frequency domain. A *Hanning* window was applied to the time domain signals to extract the surface wave components.

Three steel balls with different diameters (7, 12, and 14 mm) were used as impact sources. They generate incident surface waves with center frequencies around 25, 20, and 18 kHz. The acquired signals were digitized at a sampling frequency of 10 MHz using an NI-PXI 5105 oscilloscope.

C. Typical signal measured using air-coupled sensors

Figure 9 shows typical time domain signals measured by two air-coupled sensors. Figure 9(a) presents signals obtained from a crack-free region of the concrete specimen, where the incident waves were generated by a source with impact duration about 60 μs. The impact duration T was measured from the received time domain signals of surface waves. The direct acoustic waves were completely separated from the leaky surface waves. In Fig. 9(b), clear incident and transmitted surface waves were also obtained from the crack region (h=40 mm) of concrete. Then the surface wave components were extracted from the signals by using a Hanning window.

D. Air-coupled sensors versus accelerometers

For comparison purposes, two accelerometers (PCB model 352C65) were also installed on the concrete surface with wax, right below the air-coupled sensors at locations A and B, respectively (see Fig. 8). Figure 10 shows the surface wave transmission coefficient and signal coherence functions for signals obtained from the 40 mm deep crack region using microphones [Fig. 10(a)] and accelerometers [Fig. 10(b)].

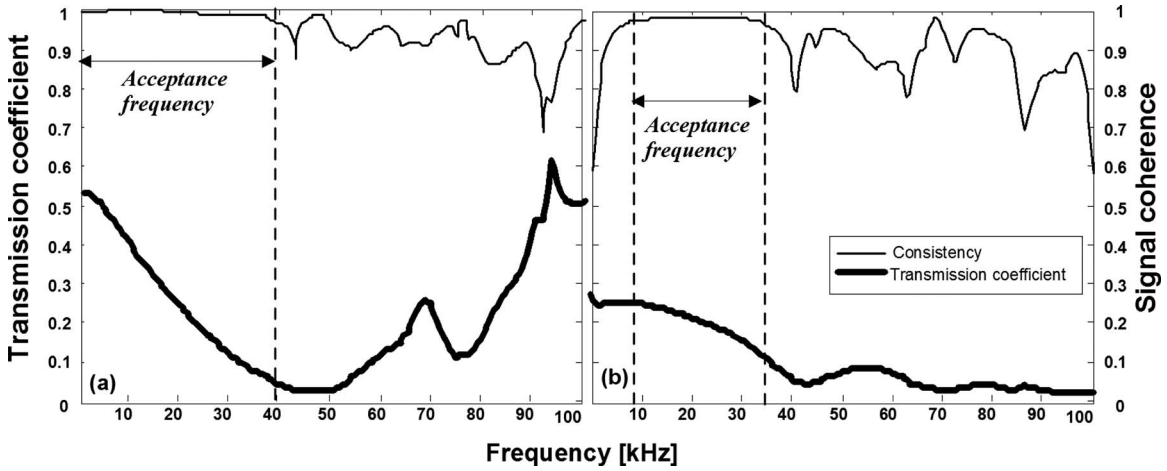


FIG. 9. Typical transmission coefficient and signal consistency versus frequency measured by (a) air-coupled sensors and (b) accelerometers. Depth of crack was 40 mm and the incident surface waves were generated by a 12 mm diameter steel ball.

The self-calibrating method was used in both tests. Both measurements show good coherence in the frequency range of 10–40 kHz. The upper limit of 40 kHz is governed by the frequency contents of the impact source. In the low frequency range (0–10 kHz); however, the air-coupled sensors show better coherence than the accelerometers. Although the accelerometers have flat frequency response in range 10–20 kHz, the temporary installation with wax causes strong resonant vibration at low frequency. This resonant vibration significantly affects signal quality around the resonant frequency, and the resonant effect cannot be removed by the self-calibrating procedure. When the center frequency of surface waves falls in this frequency range, the transmission coefficient will be unreliable. In field testing of concrete structures, the sensor coupling problem can be more critical due to rough surface of concrete and/or environmental conditions (e.g., dust, moisture, and temperature variation). Surface preparation is time consuming if large areas of concrete need to be tested. The air-coupled sensors provide a solution to this problem.

E. Calibration curve of transmission function using air-coupled sensor

The self-calibrating procedure was used to eliminate variation in the experimental setup (e.g., different impact sources, and difference between two sensors), and asymmetry in the tested specimen. Signal data were collected in the approximate far field for various crack depths with two air-coupled sensors. Three different steel balls of different diameters (7, 12, and 14 mm) were used to generate the incident surface waves. Surface wave transmission results from the three impact sources are shown in Figs. 11(a)–11(c). Overall, the experimental results show good agreement with the analytic solution, especially for shallow cracks, e.g., $h=20$ and 30 mm, and in the range of $h/\lambda<1/3$. The data corresponding to $h=10$ mm are not presented in this paper because a steel embedment under the crack distorted surface propagation. For deep cracks, the source should contain enough low frequency wave energy to give a reliable transmission measurement in $h/\lambda<1/3$ range. However, using long impact duration will induce boundary reflections from the back sur-

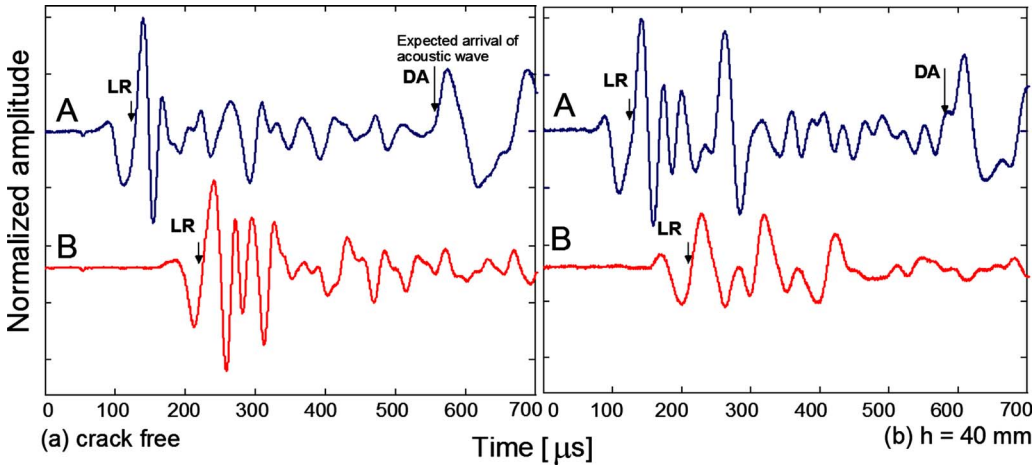


FIG. 10. (Color online) Typical signals measured by air-coupled sensors from a concrete specimen. The crack depths are (a) $h=0$ mm and (b) $h=40$ mm. The impact duration time is $T=60$ μs. Note that LR=leaky surface wave and DA=direct acoustic wave.

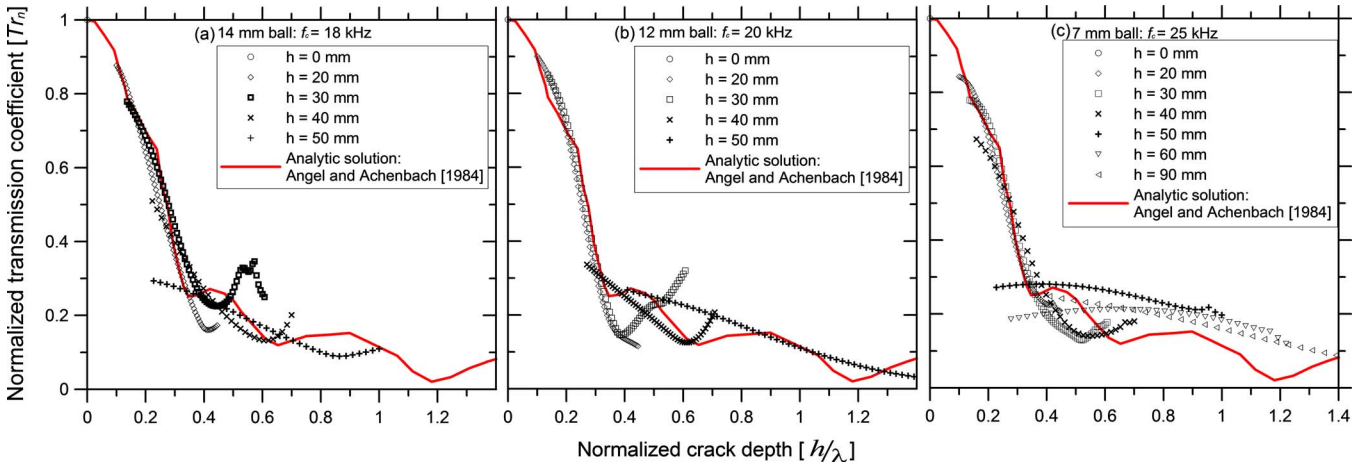


FIG. 11. (Color online) Normalized transmission coefficient and normalized crack depth relation.

face and stronger near-field effects. The crack depth that can be reliably measured using the surface wave transmission method depends on the impact source, sensor arrangement, and geometry (thickness and size) of the specimen. In this study, cracks up to 40 mm deep can be estimated with reasonable accuracy.

F. Experimental verification of the simplified algorithm using air-coupled sensor

To verify the proposed simplified algorithm for crack depth measurement, transmission coefficients measured at the center frequencies are shown in Fig. 12. The data are based on measurements at three center frequencies (18, 20, and 25 kHz) and crack depths (0, 20, 30, 40, 50, 60, and 70 mm). As shown in the figure, the experimental data are in good agreement with the analytic curve in $h/\lambda < 1/3$ range. The simplified algorithm provides a quick analysis of the measured surface wave transmission data to estimate crack depth.

IV. CONCLUSIONS

This study presents the surface wave transmission method to estimate the depth of a surface-breaking crack in concrete. First, this paper proposed a hypothesis that the near-field effect was the reason causing the discrepancy among previous research results. Numerical simulations were used to investigate the near-field effect caused by crack tip scattering and determined the near-field size. The surface wave transmission measured in the far field shows good agreement with the analytical solution, which verifies the proposed near-field effect hypothesis. Second, experimental studies were carried out on a concrete specimen to verify the numerical simulation results and demonstrate the effectiveness of the air-coupled sensing technique. The following conclusions are drawn based on the results of this study.

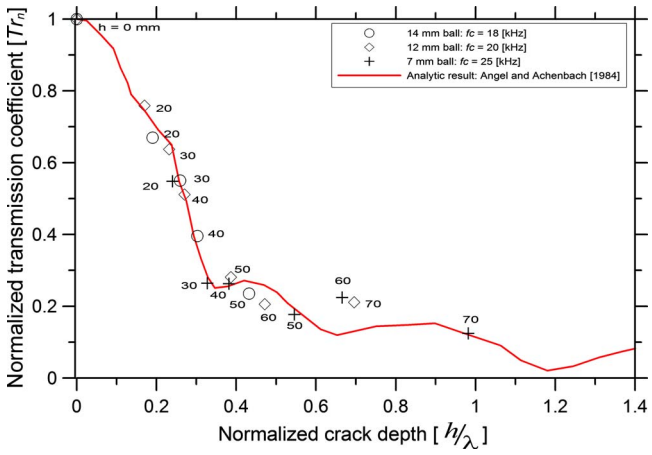


FIG. 12. (Color online) Normalized transmission coefficient based on center frequency and normalized crack depth relation.

(1) Surface wave transmission measurement is affected by near-field effects. To obtain reliable and consistent transmission coefficient, measurements should be performed in the far field. Analyses showed that the near-field size depends on the sensor location x, crack depth h, and an incident wavelength λ. An approximate near-field size [Eq. (3)] was proposed and verified effective to reduce the near-field effect in surface wave transmission measurements.

(2) The surface wave transmissions from FEM analyses and experimental studies show good agreement with the analytical curve given by Angel and Achenbach (1984). It verifies that the proposed test setup and signal processing procedure can be used to estimate of the depth of surface-breaking cracks in concrete, especially for small h/λ value ($h/\lambda < 1/3$).

(3) The air-coupled sensors show improved reliability and consistency in surface wave transmission measurements, as compared to the conventional contact sensors. Owing to the non-contact features, test speed is also improved.

(4) A simplified algorithm was proposed to estimate the crack depth based on the transmission value at the center frequency only, instead of trying to match the pre-established curve at all frequencies. If the surface wave velocity is known, the crack depth can be calculated at corresponding h/λ. The simplified algorithm has been verified by experimental data.

ABAQUS Inc. (**2007**). Analysis User's Manual, Version 6.7, Providence, RI.
Achenbach, J. D., Keer, L. M., and Mendelsohn, D. A. (**1980**). "Elastodynamic analysis of an edge crack," J. Appl. Mech. **47**, 551–556.
Achenbach, J. D., Komsky, I. N., Lee, Y. C., and Angel, Y. C. (**1992**). "Self-calibrating ultrasonic technique for crack depth measurement," J. Nondestruct. Eval. **11**, 103–108.
Alleyne, D., and Cawley, P. (**1991**). "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**, 1159–1168.
Angel, Y. C., and Achenbach, J. D. (**1984**). "Reflection and transmission of obliquely incident Rayleigh waves by a surface-breaking crack," J. Acoust. Soc. Am. **75**, 313–319.
Blackshire, J. L., and Sathish, S. (**2002**). "Near-field ultrasonic scattering from surface-breaking cracks," Appl. Phys. Lett. **80**, 3442–3444.
Cheng, A., and Achenbach, J. D. (**1996**). "A roller device to scan for surface-breaking cracks and to determine crack depth by a self-calibrating ultrasonic technique," Res. Nondestruct. Eval. **7**, 185–194.
Hevin, G., Abraham, O., Petersen, H. A., and Campillo, M. (**1998**). "Characterization of surface cracks with Rayleigh waves: A numerical model," NDT Int. **31**, 289–297.
Ihlenburg, F. (**1998**). *Finite Element Analysis of Acoustic Scattering* (Springer-Verlag, New York).
Jacobs, L. J., and Owino, J. O. (**2000**). "Effect of aggregate size on attenuation of Rayleigh surface waves in cement-based materials," J. Eng. Mech. **126**, 1124–1130.
Jian, X., Dixon, S., Guo, N., and Edwards, R. (**2007**). "Rayleigh wave interaction with surface-breaking cracks," J. Appl. Phys. **101**, 064906.
Kim, J. H., and Kwak, H.-G. (**2008**). "Nondestructive evaluation of elastic properties of concrete using simulation of surface waves," Comput. Aided Civ. Infrastruct. Eng. **23**, 611–624.
Masserey, B. and Mazza, E. (**2005**). "Analysis of the near-field ultrasonic scattering at a surface crack," J. Acoust. Soc. Am. **118**, 3585–3594.
Mendelsohn, D. A., Achenbach, J. D., and Keer, L. M. (**1980**). "Scattering of elastic waves by a surface-breaking crack," Wave Motion **2**, 277–292.
Popovics, J. S., Song, W.-J., Ghandehari, M., Subramaniam, K. V., Achenbach, J. D., and Shah, S. P. (**2000**). "Application of surface wave transmission measurements for crack depth determination in concrete," ACI Mater. J. **97**, 127–135.
Ryden, N., and Park, C. (**2004**). "Surface waves in inversely dispersive media," Near Surface Geophysics, **2**, 187–197.
Shin, S. W., Zhu, J., Min, J., and Popovics, J. S. (**2008**). "Crack depth estimation for concrete structures using spectral energy transmission of surface waves," ACI Mater. J. **105**, 510–516.
Song, W., Popovics, J. S., Aldrin, J. C., and Shah, S. P. (**2003**). "Measurement of surface wave transmission coefficient across surface-breaking cracks and notches in concrete," J. Acoust. Soc. Am. **113**, 717–725.
Viktorov, I. A. (**1967**). *Rayleigh Waves and Lamb Waves—Physical Theory and Application* (Plenum, New York).
Yew, C. H., Chen, K. G., and Wang, D. L. (**1984**). "An experimental study of interaction between surface waves and a surface breaking crack," J. Acoust. Soc. Am. **75**, 189–196.
Zhu, J., and Popovics, J. S. (**2005**). "Non-contact imaging for surface-opening cracks in concrete with air-coupled sensors," Mater. Struct. **38**, 801–806.

Viscous damping and spring force in periodic perforated planar microstructures when the Reynolds' equation cannot be applied

Dorel Homentcovschi[a] and Ronald N. Miles
Department of Mechanical Engineering, State University of New York, Binghamton, New York 13902-6000

(Received 15 July 2009; revised 13 October 2009; accepted 14 December 2009)

A model of squeeze-film behavior is developed based on Stokes' equations for viscous, compressible isothermal flows. The flow domain is an axisymmetrical, unit cell approximation of a planar, periodic, perforated microstructure. The model is developed for cases when the lubrication approximation cannot be applied. The complex force generated by vibrations of the diaphragm driving the flow has two components: the damping force and the spring force. While for large frequencies the spring force dominates, at low (acoustical) frequencies the damping force is the most important part. The analytical approach developed here yields an explicit formula for both forces. In addition, using a finite element software package, the damping force is also obtained numerically. A comparison is made between the analytic result, numerical solution, and some experimental data found in the literature, which validates the analytic formula and provides compelling arguments about its value in designing microelectomechanical devices.
 [DOI: 10.1121/1.3290990]

PACS number(s): 43.38.Bs, 43.38.Kb [AJZ] Pages: 1288–1299

I. INTRODUCTION

Recent progress in micromachining technology has enabled the fabrication of microelectromechanical systems (MEMS), such as microphones, microaccelerometers, pressure sensors, switches, mirrors, tunable interferometers, ultrasonic motors, resonators, etc. MEMS devices often use parallel plate electrodes as the capacitive sensing and electrostatic actuation mechanisms. This is why the study of a thin air layer being squeezed between a vibrating plate and a rigid plate, referred to as a planar microstructure, is important in many microelectromechanical systems. As the movable electrode displaces sinusoidally, the backforce on the plate due to the air separating it from the backplate has two components Ref. 1: the viscous damping force, which is in phase with velocity, and the spring force, which is in phase with the plate displacement. While the compressibility of the gas and inertial forces become important factors in determining the spring force at higher frequencies, the viscous forces dominate the mechanical behavior of planar microstructures at low frequencies. This paper is focused mainly on determining the viscous damping in planar, periodic perforated microstructures acting at audible frequencies.

The rigid backplate often contains numerous perforated holes which are of a few microns in diameter, uniformly distributed over the entire backplate, for reducing the time required to remove sacrificial materials between the moving structure and backplate during the wet etching process. A thicker backplate is usually preferred since it can provide much better stability and mechanical rigidity. The quasihorizontal motion of the air in the thin gap in a planar microstructure yields squeeze-film damping that can adversely affect the dynamic response of the device. Thus, in the case of microphones (and also other sensors designed for small signal applications), the mechanical-thermal noise is often one of the limiting noise components. The magnitude of thermal noise depends only on temperature and the magnitude of mechanical damping, high viscous damping being associated with large mechanical-thermal noise.[2] The perforations in one of the plates can be also used to control the viscous damping. As a result there is an extensive literature dedicated to the study of squeeze-film damping in perforated MEMS.[3–7]

While the squeeze-film damping is reduced by incorporating holes in one plate, the vertical motion of the air within the holes gives a new viscous resistance which adds to the squeeze-film damping. A rigorous solution of the total damping problem requires the solution of the Navier–Stokes' (NS) system in the three-dimensional (3D) domain comprised of the space between the plates and the volume of the holes, which is not at all a simple task. Three-dimensional flow simulations are not practical for the entire microstructure geometry[5] due to the complexity of the implementation and the computational resource requirements.

There is a large body of modeling work based on the lubrication approximation. The lubrication approximation assumes that the distance between the plane surfaces is sufficiently small, such that it is possible to integrate the NS equations over the separation distance to obtain a two-dimensional (2D) equation for pressure widely known as the Reynold's equation. In the case of perforated microstructures, the assumption that the flow domain is thin is not valid in the region of holes. As a result, the Reynold's equation

[a]Permanent address: "Politechnica" University of Bucharest and Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy, Calea13 Septembrie #13, RO-76100, Bucharest, Romania. Author to whom correspondence should be addressed. Electronic mail: homentco@binghamton.edu

cannot be applied if the air gap between the planes and the pitch and radius of holes have comparable geometrical dimensions. For this case, the widely used Reynold's equation (whose solution gives Škvor's formula[8]) may not be appropriate without significant modifications.

There are two ways to overcome the difficulties associated with the application of the Reynold's equation in perforated microstructures. The first one is to add new terms and coefficients in order to account for the influence of holes, the open end of the holes, the bending of flow from horizontal to vertical. Modifications to the Reynolds' equation have been made by adding a term related to the damping effect of gas flow through holes in Refs. 9 and 10. This modified Reynold's equation was solved analytically in particular cases, but more complicated configurations require the use of numerical methods, which diminishes the utility of the lubrication approximation. In "the analytic damping model" derived in Refs. 5, 11, and 12 (see also Ref. 13), the mechanical resistance of a perforation cell consists of six lumped flow resistances. One of these resistances is derived analytically by using the Reynolds' equation. The elongations involved in the other lumped resistances contain up to 20 numerical coefficients that are determined by using some heuristic equations and fitting to finite element method (FEM) simulations. In some cases, the new models are also encumbered with a lot of experimental coefficients added to provide agreement between the numerical and experimental results as in Refs. 14 and 15. All these methods developed for determining the damping force (and at higher frequencies also the spring force) of the oscillating perforated microstructures obtained by extending the Reynolds' equation are known as "compact models." As noted in the recent paper by Veijola *et al.*,[16] "the verification of the compact models is generally questionable."

An alternative way to determine the viscous damping of perforated microstructures when the pitch of the holes, the air gap, and the radius of holes are of the same order (when the Reynold's equation cannot be applied) is to integrate the NS system and to obtain the velocity and pressure fields. In the model developed here, we take advantage of the repetitive pattern of holes which is typical in most designs. This repetitive pattern implies the existence of a basic cell associated with each hole. For example, if the holes are aligned and regularly spaced, then the lower part of the cell (inside the gap) would be a square prism while for staggered holes, the lower part of the cell will be a hexagonal prism. Due to symmetry the normal velocity and the normal pressure derivative on each side plane of these prisms vanish. For the present model, the prismatic part of the basic cell will be approximated by a circular cylinder (of the same section area and height) having zero normal velocity along the side surface. This yields a 2D (axisymmetrical) flow problem. This approach is similar to the one used in the lubrication approximation where the original problem for a square or hexagonal domain was replaced by a problem for a circular domain which led to Škvor's formula.

The analytical solution developed in Sec. III is based on using Stokes' approximation to the Navier–Stokes system describing the motion of an isothermal, compressible fluid in the approximated cell. The resultant problem for the finite circular cylinder is similar to the flow problems studied by Robey,[17] Petritskaya,[18,19] and Zukerwar[20] for determining the theoretical response of a condenser microphone. The main result of this work is an analytical formula [Eq. (67)] for the complex force whose real part gives the viscous damping and the imaginary part is the spring force. At low frequencies (i.e., in audible domain), the viscous damping is the most important part; moreover, as was shown in Refs. 8 and 20, this component is almost independent of frequency and can be determined by a steady-state approach. This opens the possibility of determining the damping force numerically by using a finite element package for steady-state flow of compressible gas in an axisymmetrical domain.

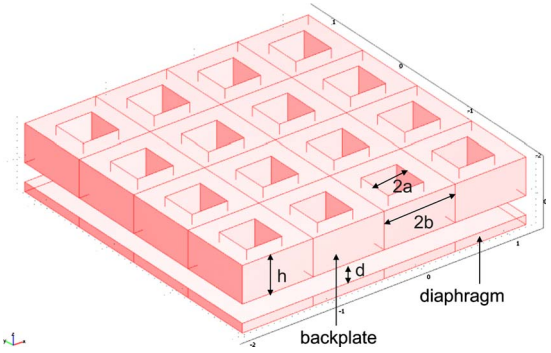


FIG. 1. (Color online) A perforated planar microstructure with aligned square holes.

The damping force given by the analytical formula is validated by results of numerical simulation and also by direct comparison with some experimental data found in the literature.

II. MODEL FORMULATION

Consider a uniform perforated (equal pitch and equal sized holes) planar microstructure having a repetitive pattern of holes. Generally, perforations are present in staggered and aligned (matrix) configurations. In the case of a staggered system of holes, a hexagonal pattern repeats while for the aligned (nonstaggered or matrix) holes, a square pattern can be noted. In Fig. 1, a regular matrix of square holes is shown. The upper plate is the perforated backplate (of thickness h) and the lower one is the diaphragm. The distance d between the average positions of the two plates is the air gap of the structure. The cylindrical holes (circular or square) lie on the vertices of a regular web of hexagons or squares of a side length denoted by $l=2b$ (the pitch of the holes) in both cases. The domain is filled by air. The geometry in the case of aligned holes (square pattern) is shown in Fig. 2. The repetitive pattern of the system of holes and the vertical motion of the diaphragm yield a similarly repetitive response in the motion of the air. The domain where a hole collects the flow will be called a cell. The basic domain for the unit cell is shown in Fig. 3. It consists of a square prism domain D_g located below the square etched cylindrical domain D_h representing the hole. All the side planes of the domain D_g are symmetry planes for the fluid motion. Consequently, on each plane surface, the normal component of velocity and the nor-

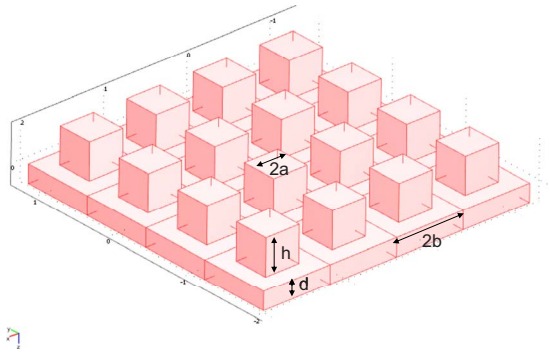


FIG. 2. (Color online) The air domain corresponding to the planar microstructure in Fig. 1.

mal pressure derivative are zero. On the other side, the velocity components vanish on the solid boundary surface of a hole (no-slip condition and no penetration conditions). Finally, the pressure on the upper base of a hole equals the external atmospheric pressure (or more generally is connected to the volume velocity by means of a known impedance Z_C of the backchamber) and the velocity on the other solid surface of the flow domain (diaphragm) equals the velocity of this surface.

In order to simplify the problem, we approximate the domain D_g by a circular cylinder C_g (see Fig. 4) of the same volume having the area of the normal section equal to the area of the hexagonal or square section of the prism D_g [the radius of the cylinder C_g is therefore $R=2b\sqrt{\sqrt{3}/(2\pi)}$ in the case of staggered holes and $R=2b/\sqrt{\pi}$ in the case of matrix holes]. The condition of vanishing of the normal velocity (and of pressure normal derivative) on all symmetry side planes will be substituted by cancellation of these quantities on the side surface of the domain C_g. The radius of the hole is denoted by r. In the case of square holes (width $2a$), as in Fig. 1, the value of r is determined by equalizing the total resistance (sum of the direct and indirect resistances from Ref. 21) for a circular hole $F^h_{\bigcirc}$ and the total resistance $F^h_{\square}$ of a square hole:

$$F^h_{\square} \equiv \frac{12}{0.4217}\mu h\left(\frac{b}{a}\right)^4 w = 8\pi\mu h\left(\frac{R}{r}\right)^4 w \equiv F^h_{\bigcirc}.$$

The resulting formula $r=1.094a$ compares well with the value $r=1.096a$ obtained in Ref. 13 by equating the acoustic

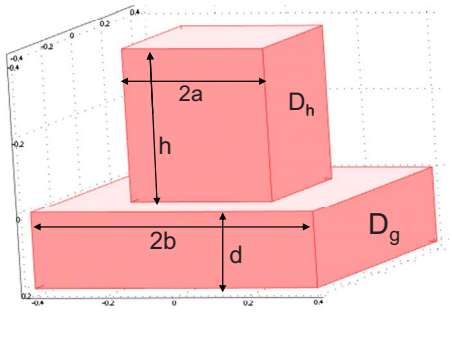


FIG. 3. (Color online) A cell of a periodic microstructure with aligned square holes.

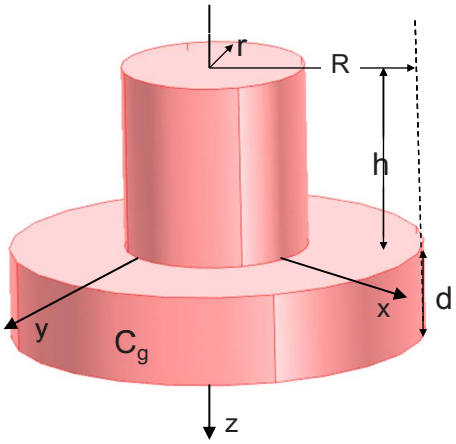


FIG. 4. (Color online) The cylindrical approximation of the basic domain (cell).

resistances (hydraulic resistances) of square and circular long channels. This approximation works in the case of small and moderate values of the area ratio A (defined as the ratio of the area of the normal section of the hole over the total area of a normal section of the domain D_g). This way we obtained an axisymmetrical problem which is easier to solve analytically and numerically.

In the case the air gap, the pitch of the holes, and the diameter of the holes are all of the same order, the lubrication approximation can no longer be used. Therefore, we no longer have a special equation for the pressure (as Reynold's equation) and in the case of isothermal flow we have to determine the complete flowfield characterized by the air velocity $\mathbf{V}$ and the pressure P. The flow inside the 3D cells is governed by the Navier–Stokes system for compressible isothermal fluid

$$\frac{\partial\tilde{\rho}}{\partial t} + \nabla\cdot(\tilde{\rho}\mathbf{V}) = 0, \tag{1}$$

$$\tilde{\rho}\left[\frac{\partial\mathbf{V}}{\partial t} + (\mathbf{V}\cdot\nabla)\right]\mathbf{V} = -\nabla P + \mu\left[\nabla^2\mathbf{V} + \frac{1}{3}\nabla\nabla\cdot\mathbf{V}\right], \tag{2}$$

where μ is the shear viscosity and $\tilde{\rho}$ is the density. Also,

$$P = c_0^2\tilde{\rho}, \tag{3}$$

c_0 denoting the isothermal speed of sound in air. The origin is placed such that the z-axis is the symmetry line of the cell in the center of the circular hole (Fig. 4). The diaphragm lies in the plane $z=d$.

III. THE ANALYTIC SOLUTION

By linearizing Eq. (2) and assuming a time harmonic solution $\mathbf{V}(\mathbf{x},t)=\mathbf{v}(\mathbf{x})\exp(-i\omega t)$, $P(\mathbf{x},t)=p(\mathbf{x})\exp(-i\omega t)$, we obtain the Stokes' system,

$$\nabla\cdot\mathbf{v} = \frac{i\omega}{\tilde{\rho}_0 c_0^2}p, \tag{4}$$

$$-i\omega\tilde{\rho}_0\mathbf{v} = -\nabla p + \mu\left(\Delta\mathbf{v} + \tfrac{1}{3}\nabla\nabla\cdot\mathbf{v}\right), \tag{5}$$

where $\omega=2\pi f$, f being the frequency and $\tilde{\rho}_0$ being the unperturbed air density. The boundary conditions on the two bases have the form

$$u(\rho,d)=0 \quad w(\rho,d)=w^{(d)}(\rho), \tag{6}$$

$$u(\rho,0)=0, \quad w(\rho,0)=w^{(0)}(\rho), \tag{7}$$

where $(u(\rho,z),0,w(\rho,z))$ are the components of velocity $\mathbf{v}$ in the cylindrical coordinates (ρ,φ,z) and the function $w^{(0)}(\rho)$ is different from zero only in the opening (the hole surface of the top of the gap).

Particularly, in the case of a small and a medium radius hole, we write

$$w^{(0)}(\rho)=\begin{cases}\tilde{w}^{(0)}(\rho) & \text{if}\rho<r\\ 0 & \text{for } r<\rho<R,\end{cases} \tag{8}$$

$\tilde{w}^{(0)}(\rho)$ being an unknown function.

Remark 1. Conditions (6) and (7) are no-slip boundary conditions assuming a continuous flow. The fluid dynamics in a narrow domain between two parallel plates yields a decrease in pressure (rarefaction effect) which changes the boundary conditions on the solid surfaces giving a certain slip (tangential velocity) of the fluid particles.[22] The presence of perforations changes the pressure distribution in the gap area next to holes. It is clear that for perforated microstructures having the pitch of holes much larger than the air gap of the microstructure, the rarefaction is still in effect. But, if the pitch of holes is not much larger than the air gap, the presence of holes makes the slip of fluid particles on solid surfaces of the microstructure at least questionable. On the other hand, an elementary calculation shows that in the case of narrow circular pipes, of infinite extent, the direct hole resistance is not affected by the slip of the fluid particles on the wall. Also, as is shown in Ref. 21, the indirect hole resistance can be considered as a particular squeeze-film damping.

A. The equations for pressure and vorticity

We write Eq. (5) as

$$\mathbf{v}=\frac{1}{i\omega\tilde{\rho}_0}\left(1-\frac{\mu}{3}\frac{i\omega}{\tilde{\rho}_0c_0^2}\right)\nabla p-\frac{\mu}{i\omega\tilde{\rho}_0}\Delta\mathbf{v}. \tag{9}$$

By using the vectorial identity,

$$\nabla\times(\nabla\times\mathbf{v})=\nabla(\nabla\cdot\mathbf{v})-\nabla\mathbf{v}, \tag{10}$$

and taking into account the equation of continuity (4), there results

$$\mathbf{v}=\frac{\beta}{i\omega\tilde{\rho}_0}\nabla p+\frac{\mu}{i\omega\tilde{\rho}_0}\nabla\times\boldsymbol{\Omega}, \tag{11}$$

where

$$\beta=1-\frac{4i\omega\mu}{3\tilde{\rho}_0c_0^2}, \tag{12}$$

and

$$\boldsymbol{\Omega}=\nabla\times\mathbf{v} \tag{13}$$

is the vorticity. In the case

$$f\ll 3\tilde{\rho}_0c_0^2/(8\pi\mu)\approx 8.95\times10^8 \text{ Hz},$$

the second term in the right-hand side of formula (12) can be neglected resulting $\beta\approx1$.

In the axisymmetrical case, in the cylindrical coordinates, there is only one nonvanishing vorticity component

$$\nabla\times\mathbf{v}=\Omega\hat{\varphi}. \tag{14}$$

Thus, we obtain expressions for the velocities in terms of the pressure p and scalar vorticity Ω as potentials,

$$\mathbf{v}=\frac{\beta}{i\omega\tilde{\rho}_0}\nabla p+\frac{\mu}{i\omega\tilde{\rho}_0}\nabla\times(\Omega\hat{\varphi}). \tag{15}$$

This relationship can also be written as

$$u(\rho,z)=\frac{\beta}{i\omega\tilde{\rho}_0}\frac{\partial p}{\partial\rho}-\frac{\mu}{i\omega\tilde{\rho}_0}\frac{\partial\Omega}{\partial z}, \tag{16}$$

$$w(\rho,z)=\frac{\beta}{i\omega\tilde{\rho}_0}\frac{\partial p}{\partial z}+\frac{\mu}{i\omega\tilde{\rho}_0}\frac{1}{\rho}\frac{\partial(\rho\Omega)}{\partial\rho}, \tag{17}$$

where $u(\rho,z)$ and $w(\rho,z)$ are the nonvanishing components of the axisymmetrical velocity field in cylindrical coordinates. By applying in formula (15), the operator $\nabla\cdot$ and accounting of relationships (4), there results the partial differential equations satisfied by pressure

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial^2p}{\partial\rho}\right)+\frac{\partial^2p}{\partial z^2}+k^2p=0, \tag{18}$$

where the scalar wave number is

$$k=\frac{\omega}{c_0\sqrt{\beta}}. \tag{19}$$

Similarly, the application of operator $\nabla\times$ and consideration of formula (14) yields the vorticity equation,

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\Omega}{\partial\rho}\right)-\frac{\Omega}{\rho^2}+\frac{\partial^2\Omega}{\partial z^2}+L^2\Omega=0, \tag{20}$$

L being the vector wave number,

$$L=\sqrt{i\omega\tilde{\rho}_0/\mu}. \tag{21}$$

B. Representation formulas for pressure, vorticity, and velocity fields

The separation of variables in Eqs. (18) and (20) associated with the condition of zero normal derivative of the pressure along the external cylindrical surface,

$$\frac{\partial p}{\partial n}(a,z)=0, \tag{22}$$

yields the representation formulas,

$$\frac{\beta}{i\omega\tilde{\rho}_0}p(\rho,z)=\phi_0(z)+\sum_{n=1}^{\infty}\phi_n(z)J_0(q_n\rho), \tag{23}$$

$$\phi_n(z) = A_0 \cos(kz) + \beta_0 \sin(kz),$$

$$\phi_n(z) = A_n \cosh(k_n z) + \beta_n \sinh(k_n z), \quad n \geq 1, \tag{24}$$

where the eigenvalues q_n are determined by

$$J_1(q_n R) = 0, \quad n = 0,1,2, \ldots, \tag{25}$$

and

$$k_n = \sqrt{q_n^2 - k^2}. \tag{26}$$

Also,

$$\frac{\mu}{i\omega\rho_0}\Omega(\rho,z) = \sum_{n=1}^{\infty} \Omega_n(z) J_1(q_n\rho), \tag{27}$$

where

$$\Omega_n(z) = C_n \cosh(r_n z) + D_n \sinh(r_n z), \tag{28}$$

and

$$r_n = \sqrt{q_n^2 - L^2}. \tag{29}$$

Formulas (16) and (17) yield the following representation formulas for the components of the velocity field:

$$u(\rho,z) = \sum_{n=1}^{\infty} u_n(z) J_1(q_n\rho), \tag{30}$$

where

$$u_n(z) = -q_n[A_n \cosh(k_n z) + B_n \sinh(k_n z)] - r_n[C_n \sinh(r_n z) + D_n \cosh(r_n z)], \tag{31}$$

and

$$w(\rho,z) = \sum_{n=0}^{\infty} w_n(z) J_0(q_n\rho). \tag{32}$$

The functions $w_n(z)$ have the expression

$$w_0(z) = k[-A_0 \sin(kz) + B_0 \cos(kz)], \tag{33}$$

$$w_n(z) = k_n[A_n \sinh(q_n z) + B_n \cosh(q_n z)] + q_n[C_n \cosh(r_n z) + D_n \sinh(r_n z)], \quad n = 1,2, \ldots. \tag{34}$$

The constants A_0, B_0, A_n, B_n, C_n, and D_n will be determined by using the boundary conditions (6) and (7).

1. Determination of constants

By using formulas (30) and (32) along the plane $z=d$, we can write

$$q_n[A_n \cosh(k_n d) + B_n \sinh(k_n d)] + r_n[C_n \sinh(r_n d) + D_n \cosh(r_n d)] = 0, \tag{35}$$

$$k_n[A_n \sinh(k_n d) + B_n \cosh(k_n d)] + q_n[C_n \cosh(r_n d) + D_n \sinh(r_n d)] = w_n^{(d)}. \tag{36}$$

Similarly, along the plane $z=0$, the u and w components of the velocity give

$$q_n A_n + r_n D_n = 0, \tag{37}$$

$$k_n B_n + q_n C_n = w_n^{(0)}, \tag{38}$$

$w_n^{(0)}$ and $w_n^{(d)}$ being the Fourier expansion coefficients corresponding to the functions $w^{(0)}(\rho)$ and $w^{(d)}(\rho)$:

$$w_n^{(0)} = \frac{2}{R^2}\int_0^r \tilde{w}^{(0)}(\rho)\rho \frac{J_0(q_n\rho)}{J_0^2(q_n R)} d\rho, \tag{39}$$

$$w_n^{(d)} = \frac{2}{R^2}\int_0^R w^{(d)}(\rho)\rho \frac{J_0(q_n\rho)}{J_0^2(q_n R)} d\rho. \tag{40}$$

Equations (35)–(38) determine the unknown pressure coefficients as

$$A_n = \mathbb{A}_n^{(0)} w_n^{(0)} + \mathbb{A}_n^{(d)} w_n^{(d)}, \tag{41}$$

$$B_n = \mathbb{B}_n^{(0)} w_n^{(0)} + \mathbb{B}_n^{(d)} w_n^{(d)}. \tag{42}$$

The coefficients $\mathbb{A}_n^{(0)}, \ldots, \mathbb{B}_n^{(d)}$ are defined by formulas

$$\mathbb{A}_0^{(0)} = \frac{\cot(kd)}{k}, \quad \mathbb{A}_0^{(d)} = \frac{\csc(kd)}{k}, \quad \mathbb{B}_0^{(0)} = \frac{1}{k}, \quad \mathbb{B}_0^{(d)} = 0, \tag{43}$$

$$\mathbb{A}_n^{(0)} = \frac{\sinh(k_n d)\cosh(r_n d) - \gamma_n \cosh(k_n d)\sinh(r_n d)}{k_n[2 - 2\cosh(k_n d)\cosh(r_n d) + [\gamma_n + \gamma_n^{-1}]\sinh(k_n d)\sinh(r_n d)]}, \tag{44}$$

$$\mathbb{A}_n^{(d)} = \frac{\gamma_n \sinh(r_n d) - \sinh(k_n d)}{k_n[2 - 2\cosh(k_n d)\cosh(r_n d) + [\gamma_n + \gamma_n^{-1}]\sinh(k_n d)\sinh(r_n d)]}, \tag{45}$$

$$\mathbb{B}_n^{(0)} = \frac{1 - \cosh(k_n d)\cosh(r_n d) + \gamma_n \sinh(k_n d)\sinh(r_n d)}{k_n[2 - 2\cosh(k_n d)\cosh(r_n d) + [\gamma_n + \gamma_n^{-1}]\sinh(k_n d)\sinh(r_n d)]}, \tag{46}$$

$$\mathbb{B}_n^{(d)} = \frac{\cosh(k_n d) - \cosh(r_n d)}{k_n[2 - 2\cosh(k_n d)\cosh(r_n d) + [\gamma_n + \gamma_n^{-1}]\sinh(k_n d)\sinh(r_n d)]}. \tag{47}$$

γ_n denotes

$$\gamma_n = \frac{k_n r_n}{q_n^2}.$$

C. One term approximations for the pressure and velocity in the opening

We assume that at the lower base of a hole the pressure has the form

$$p(\rho,0) = p_1^{(0)}, \quad \rho < r, \tag{48}$$

and that the velocity of the fluid,

$$w(\rho,0) = w^{(0)}, \quad \rho < r, \tag{49}$$

is also constant.

Particularly in the case of small holes the function $\tilde{w}^{(0)}(\rho)$ being a constant (49), a simple calculation based on formulas (A5) and (A6) yields

$$w_0^{(0)} = \frac{r^2}{R^2} w^{(0)}, \tag{50}$$

$$w_n^{(0)} = \frac{2r^2}{R^2} \frac{J_1(q_n r)}{q_n r} \frac{w^{(0)}}{J_0^2(q_n R)}, \quad n = 1,2,\ldots. \tag{51}$$

Remark 2. The one-term approximation for the pressure (48) and for velocity (49) assume small values of the radius r. For larger holes more terms have to be considered involving more unknown coefficients. However, the results obtained in Secs. IV and V show that the "one-term approximation" works well even in case of medium sized holes.

The equations of the problem are obtained now by using the representation formula (23) in condition (48):

$$\frac{\beta}{i\omega\tilde{\rho}_0} p_1^{(0)} = A_0 + \sum_{n=1}^{\infty} A_n J_0(q_n \rho), \quad \rho < r. \tag{52}$$

Equation (52) can be integrated with respect to ρ (between the limits 0 and ρ) yielding

$$A_0 \frac{\rho}{2} - \frac{\beta}{i\omega\tilde{\rho}_0} \frac{\rho}{2} p_1^{(0)} + \sum_{n=1}^{\infty} \frac{A_n}{q_n} J_1(q_n \rho) = 0, \quad \rho < r. \tag{53}$$

By substituting into Eq. (53), formulas (41) and (42), there results

$$\sum_{n=1}^{\infty} \left[\frac{\mathbb{A}_n^{(0)}}{q_n} w_n^{(0)} + \frac{\mathbb{A}_n^{(d)}}{q_n} w_n^{(d)} \right] J_1(q_n \rho) = -[\mathbb{A}_0^{(0)} w_0^{(0)} + \mathbb{A}_0^{(d)} w_0^{(d)}] \frac{\rho}{2} + \frac{\beta}{i\omega\tilde{\rho}_0} \frac{\rho}{2} p_1^{(0)}, \quad \rho < r. \tag{54}$$

Also, multiplying Eq. (54) by ρ^2 and integrating over $[0,r]$ give

$$\sum_{n=1}^{\infty} \left[\frac{\mathbb{A}_n^{(0)}}{q_n} w_n^{(0)} + \frac{\mathbb{A}_n^{(d)}}{q_n} w_n^{(d)} \right] \frac{J_2(q_n r)}{q_n r} = -[\mathbb{A}_0^{(0)} w_0^{(0)} + \mathbb{A}_0^{(d)} w_n^{(d)}] \frac{r}{8} + \frac{\beta}{i\omega\tilde{\rho}_0} \frac{r}{8} p_1^{(0)}. \tag{55}$$

Finally, the substitution of formulas (50) and (51) into relationships (55) gives the equation

$$w^{(0)} C_0 - \frac{\beta}{i\omega\tilde{\rho}_0} p_1^{(0)} = -\mathbb{A}_0^{(d)} w_0^{(d)} - \frac{8}{r} \sum_{n=1}^{\infty} \frac{\mathbb{A}_n^{(d)} J_2(q_n r)}{q_n^2 r} w_n^{(d)}, \tag{56}$$

where we have denoted

$$C_0 = \mathbb{A}_0^{(0)} \frac{r^2}{R^2} + \frac{16}{R^2 r} \sum_{n=1}^{\infty} \mathbb{A}_n^{(0)} \frac{J_1(q_n r) J_2(q_n r)}{J_0^2(q_n R) q_n^3}. \tag{57}$$

If the velocity of the diaphragm is constant, we have $w_n^{(d)} = w^{(d)} \delta_{n0}$ and the sum in right-hand side of Eq. (56) cancels out.

D. Determination of the velocity in opening

Using the impedance Z of the opening, we can write

$$p_1^{(0)} = -(Z + Z_C)\tilde{w}, \tag{58}$$

where Z_C is the acoustic impedance of the backchamber and $\tilde{w}$ is the algebraic volume velocity,

$$\tilde{\omega} = \pi r^2 w^{(0)}. \tag{59}$$

The acoustic impedance of the backchamber is[20]

$$Z_C = \frac{\gamma\tilde{\rho}_0 c^2}{i\omega V}, \tag{60}$$

where $\gamma = 1.403$ is the specific heat ratio of air and V denotes the volume of the backchamber. The impedance of the opening in the backplate is given in Ref. 23.

(1) Holes of very small diameter $r < 0.002/\sqrt{f}$,

$$Z = \frac{8\mu h}{\pi r^4} - i\omega \frac{4}{3} \frac{\tilde{\rho}_0 h}{3\pi r^2}. \tag{61}$$

(2) Intermediate-sized holes $0.01/\sqrt{f} < r < 10/f$,

$$Z \frac{\sqrt{8\omega\tilde{\rho}_0\mu}}{\pi r^2} \left(1 + \frac{h}{2r}\right) - i\omega\tilde{\rho}_0 \frac{h + 1.7r}{\pi r^2}. \tag{62}$$

Expression (62) includes also the end correction.

In the case of a hole whose radius lies between $0.002/\sqrt{f}$ and $0.01/\sqrt{f}$, interpolation must be used.

Equations (56), (58), and (59) permit determination of the pressure and velocity (assumed constants) in the opening. Particularly there results

$$w^{(0)} = W_0 w^{(d)}, \tag{63}$$

where,

$$W_0 = \left\{ k \sin(kd) \left[C_0 + \frac{Z + Z_C}{i\omega\tilde{\rho}_0} \beta \pi r^2 \right] \right\}^{-1}. \tag{64}$$

E. The analytic expressions for the damping coefficient and spring force

The pressure on the diaphragm can be written by means of formula (23) as

$$p(\rho,d) = \frac{i\omega\tilde{\rho}_0}{\beta} \left\{ A_0 \cos(kd) + B_0 \sin(kd) + \sum_{n=1}^{\infty} [A_n \cosh(k_n d) + B_n \sinh(k_n d)] J_0(q_n \rho) \right\}. \tag{65}$$

The total pressure (the pressure force) $P^{(d)}$ on the diaphragm results by integration with respect to θ and ρ:

$$P^{(d)} = -\frac{i\omega\tilde{\rho}_0}{\beta} \pi R^2 [A_0 \cos(kd) + B_0 \sin(kd)] = -\frac{i\omega\tilde{\rho}_0}{\beta} \frac{\pi R^2}{k \sin(kd)} \left[\frac{r^2}{R^2} w^{(0)} - \cos(kd) w^{(d)} \right]. \tag{66}$$

By introducing expression (63) of the velocity in the opening, we get

$$P^{(d)} = \frac{i\omega\tilde{\rho}_0}{\beta} \frac{\pi R^2 \cos(kd)}{k \sin(kd)} \left[1 - \frac{r^2}{R^2 \cos(kd)} W_0 \right] w^{(d)}.$$

Finally, the total damping coefficient of a cell B and the spring force coefficient S are

$$B + iS = \frac{i\omega\tilde{\rho}_0}{\beta} \frac{\pi R^2 \cos(kd)}{k \sin(kd)} \left[1 - \frac{r^2}{R^2 \cos(kd)} W_0 \right]. \tag{67}$$

This formula provides the total damping coefficient and the spring force coefficient on the developed model based on Stokes approximation of the viscous flow equations.

IV. AXIALLY SYMMETRIC NUMERICAL SIMULATION FOR VISCOUS DAMPING

The formula given by Škvor[8] (see also Refs. 24 and 25) assumes that the equivalent mechanical resistance of a cell (in lubrication approximation) is independent of frequency. The analysis of Zukerwar[20] shows that in the case of microphones in the audio frequency range, viscous damping is practically independent of frequency. The same conclusion was obtained in Ref. 26 in the case of an incompressible fluid as long as the frequency is smaller than 10^6 Hz. In this section, the viscous damping is modeled assuming weakly compressible fluid behavior for air. The squeezed flow is driven periodically in time and based on the above comments can be analyzed from the point of view of steady-state response.

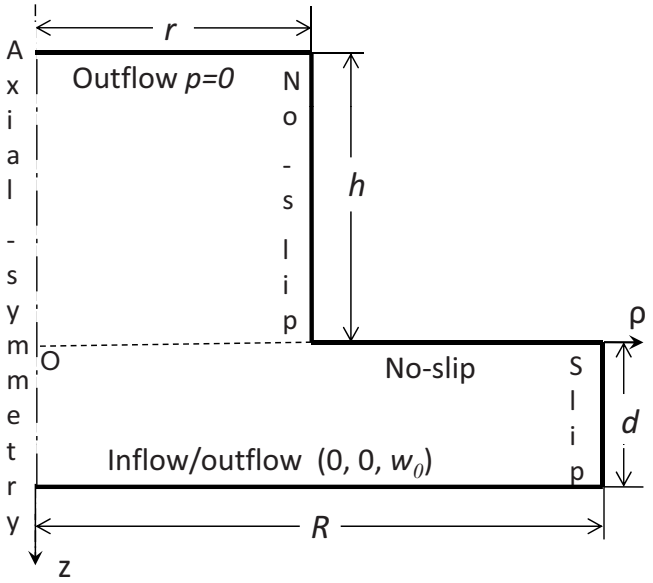


FIG. 5. The azimuthal section of the cell showing the dimensions and the boundary conditions.

The computational domain used here is the axisymmetrical region shown in Fig. 4. An azimuthal section of the cell showing the dimensions and the specified boundary conditions is given in Fig. 5. The bottom boundary corresponds to the diaphragm where the inflow/outflow condition $\mathbf{u} = (0,0,w_0)$ is specified. A constant pressure ($p=0$) outflow condition is specified along the top segment corresponding to the hole aperture. The no-slip condition is applied on solid surfaces (i.e., the wall of the hole and the lower surface of the backplate). Slip symmetry $\mathbf{V} \cdot \mathbf{n} = 0$ is specified along the outside cylindrical surface corresponding to the cell gap boundary. Finally, axisymmetry is specified along the central axis of the unit cell.

Solutions to the governing equations for the specified axisymmetric domain and boundary conditions were obtained using the weakly compressible Navier–Stokes application mode included in heat transfer module of COMSOL MULTIPHYSICS (Ref. 27) commercial software package. The software uses the finite element method for the spatial discretization of the governing equations (in cylindrical-polar coordinates). The domain is discretized using triangular elements. The COMSOL software employs an iterative procedure to solve the resulting discrete, nonlinear algebraic system of equations that result from the finite element approximation of the Navier–Stokes equations. A homogeneous, unstructured mesh was employed for most of the domain with areas of refinement in the gap region and around the re-entrant corner of the azimuthal domain. While the full Navier–Stokes equations were solved, for the conditions of the model here, the nonlinear terms are relatively small and accurate solutions are readily obtained. Grid convergence studies were performed starting with approximately 1000 elements and using up to as many as 20 000 elements for the most refined meshes.

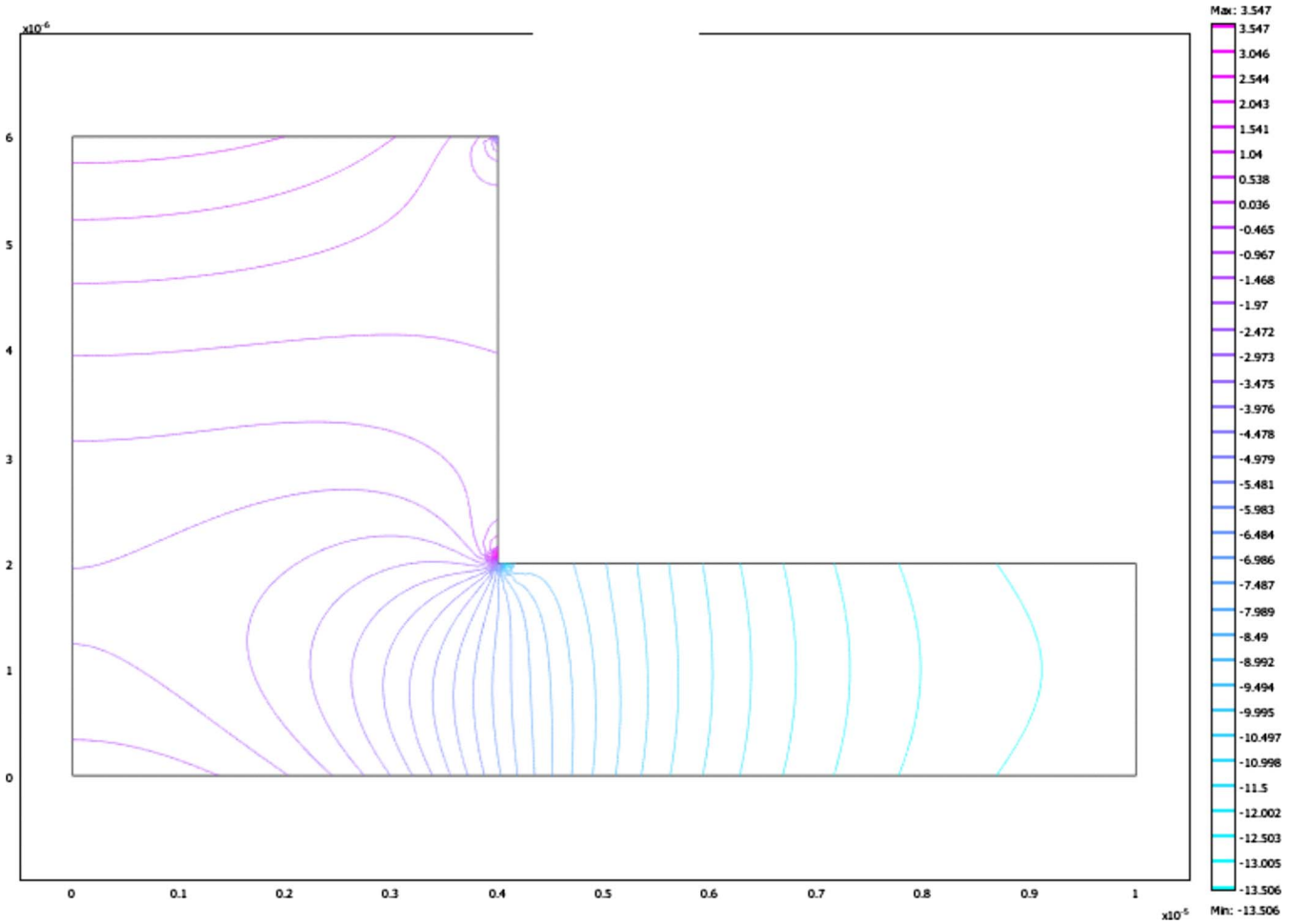


FIG. 6. (Color online) Contour plots of the pressure field in the azimuthal section resulted in numerical simulation.

Remark 3. In the numerical model presented here, standard continuum flow conditions have been assumed so that the no-slip condition applies along solid boundaries. In Ref. 16, both no-slip (models M1 and M2) and models with a slip velocity boundary condition (M3 and M4) were used along the walls in the air gap and the hole. If the scale of the flow is very small, then a slip model may be appropriate. However, for a gap on the order of 1 μm and an area ratio (area of the hole divided by the total area) around 50% half of the working area is open and rarefied gas behavior should not be important. In fact, some simple calculations show that in the case of a hole having a small radius and infinite length, the direct resistance of the hole is not influenced by the slip condition on the circular pipe. This geometry gives a redistribution of the pressure on the backplate very different from the rarefaction in a small gap between two parallel, nonperforated plane surfaces where a slip-flow may be a better model. This is evidenced by the fact that some of the results given in Table III in Ref. 16 are better when the standard continuum no-slip condition is applied along the solid surfaces. On the other hand, the indirect resistance of holes can be considered as a special squeezed film damping (see Ref. 21). See also the discussion in the book.[22]

Remark 4. In electrostatic-based transducers, application of voltages (polarization voltage for microphones and actuation voltage for actuators) is mandatory and this results in a static, nonlinear deflection of the diaphragm. In MEMS devices with very small air gap thickness, this can have a significant influence on the viscous damping. In the case of small deflections, the developed theory can be applied. For larger deflections, an axisymmetrical model can be obtained including the actual shape of the diaphragm and the FEM software will provide the viscous damping coefficient.

A. Sample results

Here we present a set of representative results by applying the axisymmetrical simulation and the analytical formula (67) to a unit cell with geometrical dimensions: $r=4$ μm, $R=8$ μm, $d=2$ μm, and $h=4$ μm. For the air in the unit cell, the following property values are used: density $\tilde{\rho}_0 = 1.155$ kg/m^3 and dynamic viscosity $\mu=18.5 \times 10^{-6}$ N s/m^2. As discussed above, a homogeneous, unstructured finite element mesh was used for the numerical computations with refinement around the re-entrant corner of the azimuthal domain. The total number of the finite elements is 9490 for the specific case presented here. Figure 6 shows contour plots of the pressure field (isobars) in the computational domain. An important element is the variation in the pressure in the axial direction (z). The result indicates

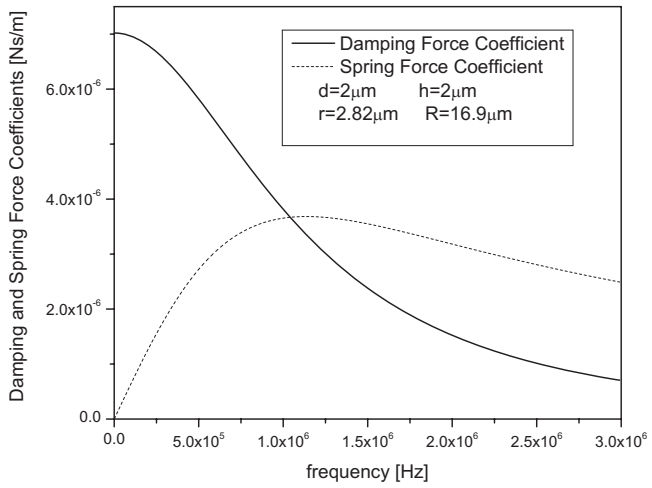


FIG. 7. Damping and spring force coefficients obtained by applying the analytical formula resulted by Stokes approximation.

a significant difference from the relationship $\partial p/\partial z=0$ assumed throughout the gap region in the lubrication type approaches.

From the numerical simulation for this case, the following values for the total damping coefficient of pressure force on the base $z=d$ of the cell (P) and the total damping coefficient of the total force (F_z) on the same base (including the viscous component) in the z-direction are obtained:

$$P = 1.058\,842 \times 10^{-7}\ \text{N s/m}, \tag{68}$$

$$B_{\text{num}} \equiv F_z = 1.058\,862 \times 10^{-7}\ \text{N s/m}. \tag{69}$$

These results show that the influence of the viscous force on the damping coefficient is very small compared with the pressure contribution. On the other hand, the calculation of the total damping coefficient as the sum of the hole resistance and the damping on the $z=0$ of the backplate gives an error of 0.23% as compared with the value given by the relationship (69).

Analytical formula (67) yields the following value of the damping coefficient:

$$B_{\text{STO}} = 1.1356 \times 10^{-7}\ \text{N s/m}, \tag{70}$$

resulting in a relative error of $e_{\text{STO}}=6.6\%$ with respect the value given by FEM simulation. The damping coefficient corresponding to the Reynolds approximation is

$$B_{\text{Rey}} = 6.15 \times 10^{-8}\ \text{N s/m}. \tag{71}$$

This value was obtained by adding the squeezed film damping given by Škvor's formula and the direct and indirect resistances of a hole defined in Ref. 21. The resulting value B_{Rey} has a relative error with respect to the FEM value of $e_{\text{Rey}}=42\%$.

The calculation of the total force on a cell by using the obtained analytical formula (67) was performed in the range of frequencies from 100 Hz to 3 MHz. The results are shown in Fig. 7, where the continuous line represents the damping force and the dotted line represents the spring force in the case where the effects of inertia and compressibility were included. The damping force dominates at low frequencies. The cutoff frequency is ≈ 1.5 MHz. For higher frequencies, the spring force is more important.

V. VALIDATION OF THE ANALYTICAL FORMULA

In this section, we compare the predictions from analytical solution and the FEM model to the published experimental data of Somà and De Pasquale.[28] In this cited paper, the details of the device configuration studied, the experimental setup, and the testing procedure are presented. The experimental configuration consists of a central suspended plate (where the structural stiffness is concentrated) covered with a pattern of square holes and connected to four lateral clamped supports of small cross-sectional area. This particular shape allows an out-of-plane deflection of lateral supports during the excitation and, consequently, a quasirigid oscillation of the central plate. The suspended plates are $h=15$ μm thick and have a gap of $d=1.6$ μm. Six different test configurations were studied: four specimens with different hole cross-sectional area size and two with the same hole dimensions but different plate widths. Table I gives the effective geometrical dimensions of the test structures obtained by profile measurements using an interferometric microscope. In the table, six cases are labeled A–F as in Refs. 16 and 28. The first four columns in Table I coincide with the corresponding columns in Table 1 of Ref. 16 including the plate length L, the plate width W, and the number of holes $M \times N$. Column five contains the perforation ratio q (same as the area ratio A) of the structure. In columns six and seven of Table I, we include the parameters b (the half-period of the structure) and $a=s_0/2$ (the half-length of a square hole side). Column nine contains the radius r of the circular cylinder which is equivalent to the square cross section of the hole of the test

TABLE I. The geometry of the measured and calculated structures.

Type	L (μm)	W (μm)	$M \times N$	q (%)	b (μm)	a (μm)	r (μm)	R (μm)
A	372.4	66.4	36×6	24	5.1	2.5	2.73	5.75
B	363.9	63.9	36×6	37	5.0	3.05	3.34	5.64
C	373.8	64.8	36×6	50	5.15	3.65	4	5.81
D	369.5	64.5	36×6	59	5.1	3.95	4.23	5.75
E	363.8	123.8	36×12	38	5.0	3.1	3.39	5.64
F	363.8	243.8	36×24	38	5.0	3.1	3.39	5.64

TABLE II. The measured, the analytical (Reynolds and Stokes) and computed damping coefficients.

Type	C_M measured (10^{-6} N s/m)	C_{Rey}-ΔRey (10^{-6} N s/m)-(%)	C_{STO}-ΔSTO (10^{-6} N s/m)-(%)	C_{num}-Δnum (10^{-6} N s/m)-(%)
A	47.38	33.82(−28.60%)	40.81(−13.88%)	40.46(−14.61%)
B	19.46	14.00(−28.10%)	18.30(−5.98%)	17.84(−8.32%)
C	9.863	7.57(−23.25%)	10.76(9.14%)	10.20(3.42%)
D	7.609	5.65(−25.81%)	8.095(6.39%)	7.650(0.54%)
E	38.22	26.35(−31.06%)	34.66(−9.30%)	33.74(−11.7%)
F	67.44	52.70(−21.86%)	69.32(2.89%)	67.48(0.06%)

specimen. Finally, the last column contains the radius R of the cylinder which substitutes the lower prism (gap region) of the domain.

In Ref. 28, the quality factor is determined from the experimentally obtained curve of displacement versus frequency, and the damping coefficient is then calculated from the quality factor, frequency, and the effective mass using the half-power bandwidth method. The values of the measured total damping coefficients, labeled as C_m, are given in column 2 of Table II for all six cases studied (note that this coincides with the second column in Table 3 of Ref. 28).

The columns 3–5 of Table II here contain the total damping coefficients C_{Rey}, C_{STO}, and C_{num} for the plate determined by multiplying the damping coefficient for a unit cell with the total number $M \times N$ of holes. In parenthesis in each case the relative error (ΔRey, ΔSTO, and Δnum) of the respective damping coefficient with respect to the measured value C_M was included. The coefficient c_{Rey} for a single cell was obtained by adding the squeezed film damping given by Škvor's formula and the direct and indirect resistances of a hole defined in Refs. 21 and 26. The coefficient c_{STO} is given by the real part of formula (67), while the coefficient c_{num} is obtained in numerical simulation as described in Sec. IV A.

It is clear that the results given by analytical formula are comparable with those obtained by finite element simulation and close to the measured values. Particularly, this validates directly the formula (67) for determining the damping and spring force for a cell. The precision of the analytical results are less in the case of perforated microstructures having a high area ratio. A better analytical approach in this case needs a more precise representation of the velocity in the opening. Also the total damping coefficient can be corrected by adding an edge correction to the boundary cells similar to the correction introduced in Ref. 21 for the case of the Reynolds' equation.

The results included in column 3 in Table II show that the Reynolds' lubrication model predictions fall below the results from the other two models. In some cases even the first digit of the total damping coefficient C_{Rey} is incorrect.

Table III presents the relative errors in the damping coefficient between the experimental values reported in Ref. 28, the four compact models presented in Ref. 16, and the relative error from the methods developed in this paper. The second column in Table III gives the relative error ΔM1 corresponding to the compact model M1 of Bao[10] for a rectangular damper that has a much larger length than width. The relative error values in column 3 (ΔM2) correspond to the model presented in Ref. 10 for an arbitrary rectangular surface. Columns 4 and 5 contain the relative errors (ΔM3 and ΔM4) of the compact models developed by Veijola:[5] model M3 corresponds to circular perforations of the backplate while model M4 was tailored specifically for square holes.[12] Column 6 includes the relative errors ΔSTO resulting from the use of the presented analytical formula and the last column includes the errors Δnum of the numerical simulation by the finite element method.

The results of numerical simulation given in this paper are in closer agreement with the measured values than those of the compact methods for all the microstructures considered. Also the values for the damping coefficient obtained by using the analytical formula for the damping coefficient of a cell compares well with the results reported in Ref. 16 for the compact models.

VI. CONCLUSIONS

A model for analyzing the viscous damping in a unit cell of a perforated microstructure has been developed. The model is based on the Stokes' equations for compressible, isothermal, viscous flow, and the approximation of the peri-

TABLE III. Relative errors of the compact models of analytical formula and of numerical simulation. The values ΔM1 – ΔM4 were given in Ref. 16.

Type	ΔM1 (%)	ΔM2 (%)	ΔM3 (%)	ΔM4 (%)	ΔSTO (%)	Δnum (%)
A	−23.53	−25.74	−33.51	−33.27	−13.88	−14.61
B	−16.36	−18.06	−21.02	−21.96	−5.98	−8.32
C	−5.21	−6.59	−4.11	−6.65	9.10	3.42
D	−14.66	−15.72	−12.46	−15.29	6.39	0.54
E	−17.27	−18.94	−19.03	−20.14	−9.30	−11.7
F	−4.77	−6.70	−5.19	−6.52	2.89	0.06

odic structure of common devices (squares or hexagonal prisms) by an equivalent axisymmetrical cylindrical domain. The analysis yields an analytical formula for the complex force (damping force and spring force) generated by the vibrating plate of the microstructure.

In addition, a FEM solution was used to determine the damping on the cell in the axisymmetrical domain.

The predictions for the damping coefficients from the analytical formula and numerical simulation were compared for some test microstructures and are found to be in good agreement. Also, these results are found to be in good agreement with some measured values found in the literature for some MEMS devices. This validates the model and demonstrates its broader range of applicability to real perforated microstructures. Also, the results obtained by these models compare well with the values given by some compact models in appropriate parameter ranges.

ACKNOWLEDGMENTS

This work has been supported by the National Institute on Deafness and Other Communication Disorders under Grant No. R01DC005762-05 for a NIH Bioengineering Research Partnership and Grant No. R01DC009429 to R.N.M.

APPENDIX: FINITE HANKEL TRANSFORMS AND MODIFIED FINITE HANKEL TRANSFORMS

In this appendix we prove the relationships (A5) and (A6) which are used in Sec. III A.

1. Some definitions and properties

We consider the finite Hankel transform of the function $u(\rho)$ with respect to ρ-variable defined by the relationship[29] (p. 83)

$$\tilde{u}(q_i) \equiv \tilde{\mathbb{J}}_1[u(\rho)] = \int_0^R u(\rho)\rho J_1(q_i\rho)d\rho, \tag{A1}$$

where q_i is a root of the transcendental equation

$$J_1(q_iR) = 0. \tag{A2}$$

At any point of $(0,R)$ at which the function $u(\rho)$ is continuous, the original function can be recovered by means of the inversion formula

$$u(\rho) = \frac{2}{R^2}\sum_i \tilde{u}(q_i)\frac{J_1(q_i\rho)}{[J_0(q_iR)]^2}, \tag{A3}$$

where the sum is taken over all the positive roots of Eq. (A2).

We define also the modified finite Hankel transform of the function $w(\rho)$ by formula

$$\hat{w}(q_i) \equiv \hat{\mathbb{J}}_0[w(\rho)] = \int_0^R w(\rho)\rho J_0(q_i\rho)d\rho. \tag{A4}$$

Inversion formula for the modified finite Hankel transform can be written as

$$w(\rho) = \frac{2}{R^2}\sum_{i=0}^{\infty} \hat{w}(q_i)\frac{J_0(q_i\rho)}{[J_0(q_iR)]^2}.$$

2. Calculation of some transforms

In the case of the function

$$f(\rho,r) = \begin{cases} \rho, & 0 \le \rho < r \\ 0, & r < \rho < R, \end{cases}$$

there results

$$\tilde{f}(q_ir) = \int_0^r \rho^2 J_1(q_i\rho)d\rho = r^3\int_0^1 t^2 J_1(q_irt)dt.$$

The last integral is given again by formula (6.567(1)) in Ref. 30. Therefore,

$$\tilde{f}(q_i,r) = r^3\frac{J_2(q_ir)}{q_ir}. \tag{A5}$$

In the case of the function,

$$h(\rho,r) = \begin{cases} 1, & 0 \le \rho < r \\ 0, & r < \rho < R, \end{cases}$$

there results

$$\hat{h}(q_i,r) = \int_0^r \rho J_0(q_i\rho)d\rho = r^3\int_0^1 tJ_0(q_irt)dt.$$

The last integral is given by formula (6.567(9)) in Ref. 30,

$$\hat{h}(q_i,r) = r^2\frac{J_1(q_ir)}{q_ir}. \tag{A6}$$

[1] J. J. Blech, "On isothermal squeeze films," J. Lubr. Technol. **105**, 615–620 (1983).
[2] T. B. Gabrielson, "Mechanical-thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**, 903–909 (1993).
[3] J. B. Starr, "Squeeze-film damping in solid-state accelerometers," in Proceedings of the IEEE Solid State Sensor and Actuator Workshop, Hilton Head Island, SC (1990), pp. 44–47.
[4] Y. J. Yang and S. D. Senturia, "Numerical simulation of compressible squeezed-film damping," in Proceedings of the IEEE Solid State Sensor and Actuator Workshop, Hilton Head Island, SC (1990), pp. 76–79.
[5] T. Veijola, "Analytic damping model for an MEM perforation cell," Microfluid. Nanofluid. **2**, 249–260 (2006).
[6] M. Bao and H. Yang, "Squeeze film air damping in MEMS," Sens. Actuators, A **136**, 3–27 (2007).
[7] C. W. Tan, Z. Wang, J. Miao, and X. Chen, "A study on the viscous damping effect for diaphragm-based acoustic MEMS applications," J. Micromech. Microeng. **17**, 2253–2263 (2007).
[8] Z. Škvor, "On acoustical resistance due to viscous losses in the air gap of electrostatic transducers," Acustica **19**, 295–297 (1967).
[9] T. Veijola and T. Mattila, "Compact squeezed-film damping model for perforated surface," in Proceedings of the Transducers '01, München, Germany (2001), pp. 1506–1509.
[10] M. Bao, H. Yang, Y. Sun, and P. J. French, "Modified Reynolds' equation and analytical analysis of squeeze-film air damping of perforated structures," J. Micromech. Microeng. **13**, 795–800 (2003).
[11] S. S. Mohite, V. R. Sonti, and R. Pratap, "A compact squeeze-film model including inertia, compressibility, and rarefaction effects for perforated 3-D MEMS Structures," J. Microelectromech. Syst. **17**, 709–723 (2008).
[12] T. Veijola, "Analytic damping model for a square perforation cell," in Proceedings of the Ninth International Conference on Modeling and Simulation of Microsystems, Boston (2006), Vol. **3**, pp 554–557.
[13] T. Veijola and P. Raback, "Methods for solving gas damping problems in

perforated microstuctures using a 2D finite-element solver," Sensors **7**, 1069–1090 (2007).
[14]S. S. Mohite, V. H. Kesari, V. R. Sonti, and R. Pratap, "Analytical solutions for the stiffness and damping coefficients of squeeze films in MEMS devices having perforated back plates," J. Microelectromech. Syst. **15**, 2083–2092 (2005).
[15]A. K. Pandey and R. Pratap, "A comparative study of analytical squeeze film damping models in rigid rectangular perforated MEMS structures with experimental results," Microfluid. Nanofluid. **4**, 205–218 (2008).
[16]T. Veijola, G. De Pasquale, and A. Somà, "Comparison between damping coefficients of measured perforated structures and compact models," in Proceedings of the DTIP, Nice (2008), pp. 236–241.
[17]D. H. Robey, "Theory of the effect of a thin air film on the vibrations of a stretched circular membrane," J. Acoust. Soc. Am. **26**, 740–45 (1954).
[18]I. G. Petritskaya, "Impedance of a thin layer of air in the harmonic vibrations of a membrane," Sov. Phys. Acoust. **12**, 193–198 (1966).
[19]I. G. Petritskaya, "Vibrations of a membrane loaded with a thin layer of air," Sov. Phys. Acoust. **14**, 105–106 (1968).
[20]A. J. Zuckerwar, "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285 (1978).
[21]D. Homentcovschi and R. N. Miles, "Viscous microstructural dampers with aligned holes: Design procedure including the edge correction," J. Acoust. Soc. Am. **122**, 1556–1567 (2007).
[22]G. Karniadakis, A. Beskok, and N. Aluru, *Microflows and Nanoflows* (Springer, Berlin, 2005).
[23]L. L. Beranek, *Acoustics* (McGraw-Hill, New York, 1954), p. 137.
[24]D. Homentcovschi and R. N. Miles, "Modelling of viscous damping of perforated planar micro-mechanical structures. Applications in acoustics," J. Acoust. Soc. Am. **116**, 2939–2947 (2004).
[25]D. Homentcovschi and R. N. Miles, "Viscous damping of perforated planar micromechanical structures," Sens. Actuators, A **119**, 544–552 (2005).
[26]D. Homentcovschi and R. N. Miles, "Analytical model for viscous damping and the spring force for perforated planar microstructures acting at both audible and ultrasonic frequencies," J. Acoust. Soc. Am. **124**, 175–181 (2008).
[27]COMSOL MULTIPHYSICS Version 3.5, http://www.comsol.com. (Last viewed 6/22/2009).
[28]A. Somà and G. De Pasquale, "Identification of test structures for reduced order modeling of the squeeze film damping in MEMS," Proceedings of the DTIP Symposium on Design, Test, Integration and Packaging of MEMS & MOEMS (2007), pp. 230–239.
[29]I. Sneddon, *Fourier Transforms* (McGraw-Hill, New York, 1951).
[30]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1994).

Analysis of the attenuation of railway squeal noise by preloaded rings inserted in wheels

J. F. Brunel,[a] P. Dufrénoy, and J. Charley
Laboratoire de Mécanique de Lille CNRS UMR 8107, Cité Scientifique, 59655 Villeneuve d'Ascq, France

F. Demilly
Valdunes, BP 12, 59125 Trith Saint Léger, France

(Received 22 January 2009; revised 20 October 2009; accepted 31 December 2009)

Squeal from railway wheels occurring in short radius curves produces a very intense and highly annoying noise in the range 400–8000 Hz. When the excitation, due to lateral forces acting on the wheel, cannot be avoided, additional systems can be added on the wheel to limit acoustic emission. A very economical approach is the use of metal rings inserted into grooves machined in the wheels. Unfortunately the effectiveness of these so called *damping rings* varies from one wheel to another and for different rings. Because the mechanisms of attenuation are not well understood, these variations have not to date been explained. The aim of this paper is to clarify the attenuation mechanisms for damping rings especially for the first three axial wheel modes, which are the predominant sound radiated ones in curve passage and for which the effectiveness of the treatment is lower. It has been generally assumed that friction between the ring and the groove has been the mechanism for squeal noise attenuation. Here it is shown that the vibration attenuation is due to modal coupling between the wheel and the ring. The validity of this proposed mechanism is investigated using experimental measurements and theoretical and numerical models. The results presented here will provide an avenue for optimization of the damping ring noise control treatment to obtain significant levels of squeal noise attenuation notably for the first three axial modes.
 [DOI: 10.1121/1.3298433]

PACS number(s): 43.40.At, 43.50.Lj [ADP] Pages: 1300–1306

I. INTRODUCTION

Curve squeal is one of the most significant source of noise annoyance on metros and tramways. Noise levels can reach 140 dB close to the wheel. Squeal noise is due primarily to the excitation of one or more axial wheel modes from 400 Hz to 10 kHz (Vincent *et al.*, 2004). The origin of curve squeal was first explained by Rudd (Rudd, 1976) He concluded that in a curve in addition to rolling the wheel slides laterally across the rail, resulting in lateral contact forces that lead to self-excited vibrations of the wheel (Brunel and Dufrénoy, 2007). A great number of wheel modes are then excited leading to acoustic emissions with overall sound pressure levels up to 120 dB close to the wheel (Thompson and Jones, 2000). At macroscopic scale, friction forces depend on various parameters: curve radius, wheel modal frequencies, and damping of the wheel. At the microscopic scale, the self-excited vibrations are generated by a feedback mechanism involving dry friction and the motion of the wheel relative to the rail. Parameters are average lateral velocity, contact pressure distribution, microscopic properties of the wheel, and rail surfaces. Although many parameters strongly influence the excitation (and hence the occurrence and intensity) of squeal, it is proposed in this work to use a simplified excitation, consisting of a harmonic lateral force on the wheel as would be induced by the lateral sliding of the wheel on the rail as a rail vehicle rounds a curve (Rudd, 1976). This simplified approach is justified since the objective of this work is understanding the mechanisms of a specific acoustic attenuation system. It is also well known by experiments that the modes excited during squealing in curves are very close to the eigenmodes of the wheel with nodal lines consisting of one circle and several diameters (Wetta and Demilly, 1985). Figure 1(a) illustrates this point where the peaks correspond to n-m modes of the wheel with $n=0$ (the number of circles) and $m=2-5$ (the number of lines or diameters). These modes will, of course, be well excited by the proposed excitation.

Various treatments for the mitigation of curve squeal noise have been developed. The use of damping rings is one of the most cost effective. Viscoelastic treatments or acoustic baffles are typically much more expensive. To employ the treatment, wheels are equipped with one or two rings inserted into grooves in the rim of the wheel. Acoustic measurements on trains under normal operating conditions show that these ring-damped wheels can reduce squeal noise by about 10 dB at a distance of about 7 m from the track [Fig. 1(b)] (Wetta and Demilly, 1985) when compared to existing wheels. However, it has been observed that the effectiveness depends on the application and may vary a great deal from one wheel to another with no plausible explanation. An illustration of this point is presented in the second part of this paper. Reliable performance of the ring damping treatment can only be achieved, if the physical mechanism associated with the noise attenuation is well understood. While friction or impact between the ring and groove are usually thought to

[a]Author to whom correspondence should be addressed. Electronic mail: jean-francois.brunel@polytech-lille.fr

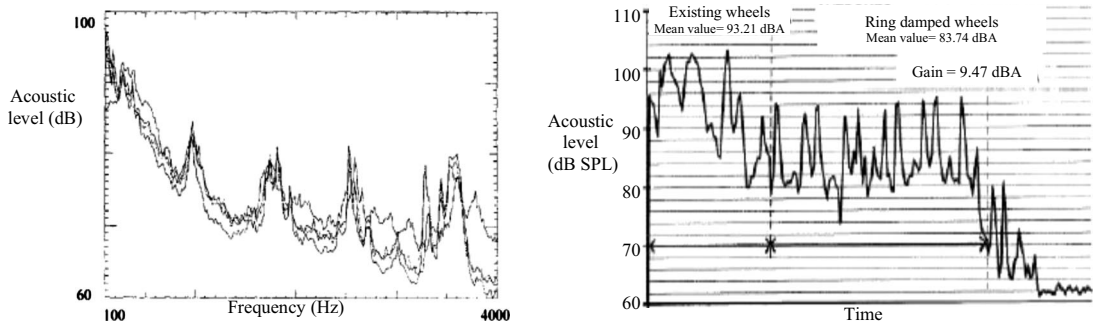


FIG. 1. (a) Squeal noise spectra measured on "Corail" wheels. (b) Noise level time history for undamped and ring-damped wheel.

be the mechanism for squeal noise attenuation (Akay, 2002; López, 1999; López *et al.*, 2004), recent numerical investigations have shown that these two mechanisms can only account for a reduction of about 1 dB in wheel vibration and noise (Brunel *et al.*, 2006), not the normally observed 10 dB, mentioned above.

Experimental measurements on the wheel with a ring have been performed and are presented in Sec. III where the effect of tightness of the ring in the groove is examined. Those results are discussed and used for comparison with the theoretical approach developed in Sec. IV. From these results, an original explanation of the mechanism of squeal noise attenuation by damping rings is proposed. A three-dimensional (3D) numerical model of the railway wheel is employed in Sec. V to validate the proposed explanation. Finally, conclusions and perspectives are discussed in Sec. VI.

II. DESCRIPTION AND EFFECTIVENESS OF RING DAMPERS

The damping ring treatment is a simple way to increase damping of axial modes in the wheel. To apply the treatment, one or two metal rings are inserted into grooves machined into the rim of the wheel [Fig. 2(a)]. The ring is first inserted into the groove by compressing it circumferentially. Afterwards it is welded together at its open end [Fig. 2(b)]. Attenuation is about 10 dB [as shown in Fig. 1(b)] and can reach up to 20 dB for particular conditions. Several conclusions may be summarized from previous experimental investigations (Brunel *et al.*, 2003). The noise reduction depends directly on the assembly conditions and particularly on the welding operation. The ring insertion leads to a damping increase (mean loss factors are as high as 0.3% in the frequency range of 3000–8000 Hz for the double damping ring treatment compared to 0.07% for the untreated wheel). The loss factors are increased primarily in the first three eigenmodes excited during squealing (0-2, 0-3, 0-4). For a single ring, the loss factors are around 0.12%. Significant variations in the noise reduction performance are obtained even for similar wheels with rings welded under the same conditions.

Preload between the ring and the groove seems to be the most important factor in determining the noise attenuation (in industrial conditions, preload is obtained at the welding operation but is not controlled during the process). This factor is controlled in this study by equipping a 25 mm diameter ring with a screw-nut system inserted between the two ends of the ring [Fig. 2(c)]. The preload is measured by strain gauges on the ring and by the relative displacements of the ring's extremities (Brunel *et al.*, 2004b).

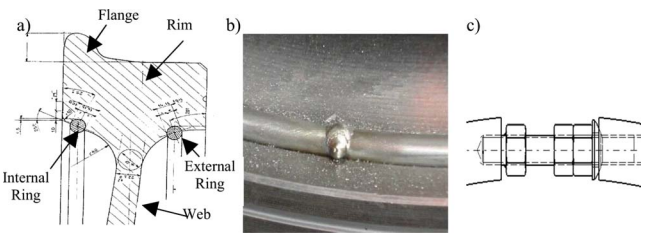


FIG. 2. (Color online) The ring-damped wheel and the system for controlling preload.

III. EXPERIMENTAL INVESTIGATION OF THE ATTENUATION MECHANISMS

A. Experimental setup

Experimental measurements were performed in a reverberant room on a wheel equipped with an external ring [Fig. 2(a)]. The focus was on one squealing mode corresponding to the 0-3 mode of the wheel, at around 1100 Hz (Thompson and Jones, 2000) (Brunel *et al.*, 2004a). The wheel was excited in the axial direction with a dynamic shaker (Fig. 3) driven with white noise (0–8000 Hz). The purpose of the shaker was to simulate a lateral force excitation as explained in the Introduction. That simulation is, of course, not perfect in that the force level is relatively low (<100 N) and that the feedback mechanism between the dry-friction force in the wheel/rail contact and the motion relative to the rail cannot be simulated. However, results show that the excited fre-

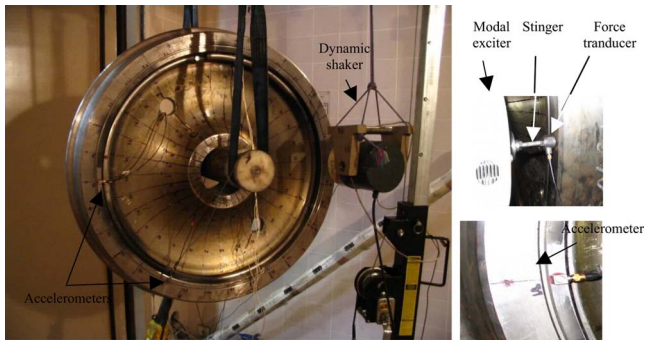


FIG. 3. (Color online) Experimental setup.

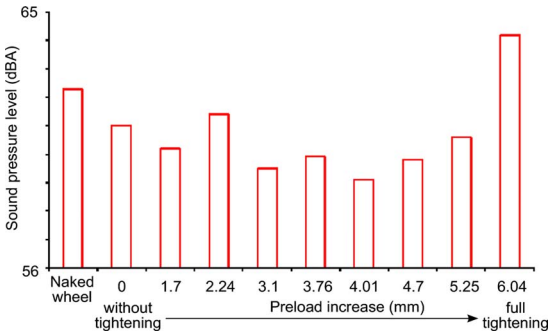


FIG. 4. (Color online) Sound pressure level vs preload.

quencies are the same as those measured near a curved track during a train passage and that the radiated noise levels are significant and concern the right frequencies.

The global noise level in the reverberant room has been determined in the range 500–4000 Hz and is plotted in Fig. 4 with various ring preload values. Values of ring preload refer to the distance between the two extremities of the ring. Note that the case "without tightening" corresponds to the case where the ring was inserted into the groove without expanding the ring to create a supplementary force.

Results show that increasing the preload first decreases the sound pressure level until a middle range of tightening with minimum noise level. For higher preload, the noise level increases to values greater than achieved with the naked wheel. Note that for the case without tightening, some acoustic attenuation exists due to the initial preload obtained during assembly. Several configurations have been tested showing the repeatability of this observation with a middle range of preload for which acoustic emissions are lower.

B. Acceleration spectra

In addition to acoustic measurements, the acceleration of the wheel and the ring were measured by accelerometers fixed on the wheel and on the ring at the same angular position. Waterfall plots of the axial wheel acceleration vs ring preload are shown in Fig. 4(a). Each spectrum of the waterfall plots of Fig. 5(a) corresponds to a preload value of the ring. The last spectrum relates to the naked wheel response. Note that the successive frequency peaks correspond to eigenmodes of the wheel as seen on the naked wheel spectrum and that due to the direction of excitation, these modes are 0-*n* type corresponding to those excited during squeal. For high frequencies (greater than 3000 Hz), the ring damper seems to be effective for a large preload band. For the first four modes, comparisons with the naked wheel show that the reduction in wheel acceleration depends strongly on the preload and that there is an "optimal range" of preload where the acceleration is minimized.

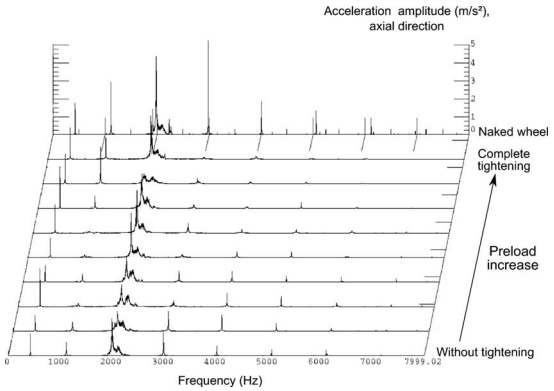


FIG. 5. Waterfall plots of the wheel acceleration vs preload and frequency.

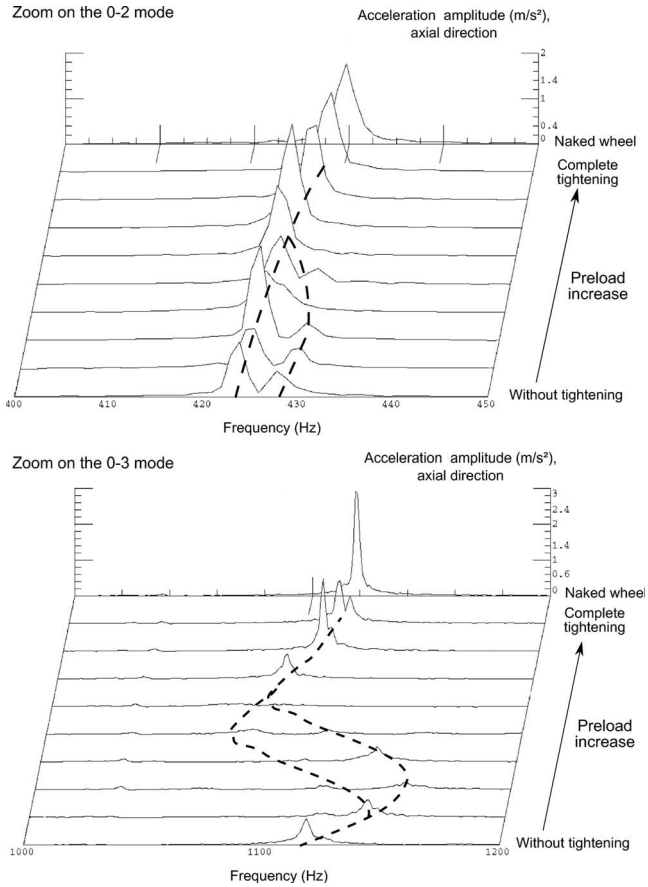


FIG. 6. Evolution of the wheel modes with preload increase.

The following highlights some of the observations with regard to the effect of ring preload.

- Slight variation of the first two modal frequencies with preload (see Fig. 6). The range of preload where the acceleration is significantly reduced is the same as the range where the modal frequencies change the most. As the preload is increased, the wheel modal frequencies first increase and then rapidly decrease. The frequencies then increase but stay below the modal frequencies of the naked wheel.
- Decrease in the acceleration levels of the wheel and ring (see Fig. 7).

In the experiments, the ring-damped wheel is excited with a sinusoidal function at the exact frequency of the wheel eigenmode (adjusted for each preload). Figure 6 presents the axial acceleration spectra of the wheel and the ring vs preload for a sinusoidal excitation of the 0-3 mode of the wheel at approximately 1100 Hz. Around the optimal preload, attenuation of the vibration of both the wheel and the

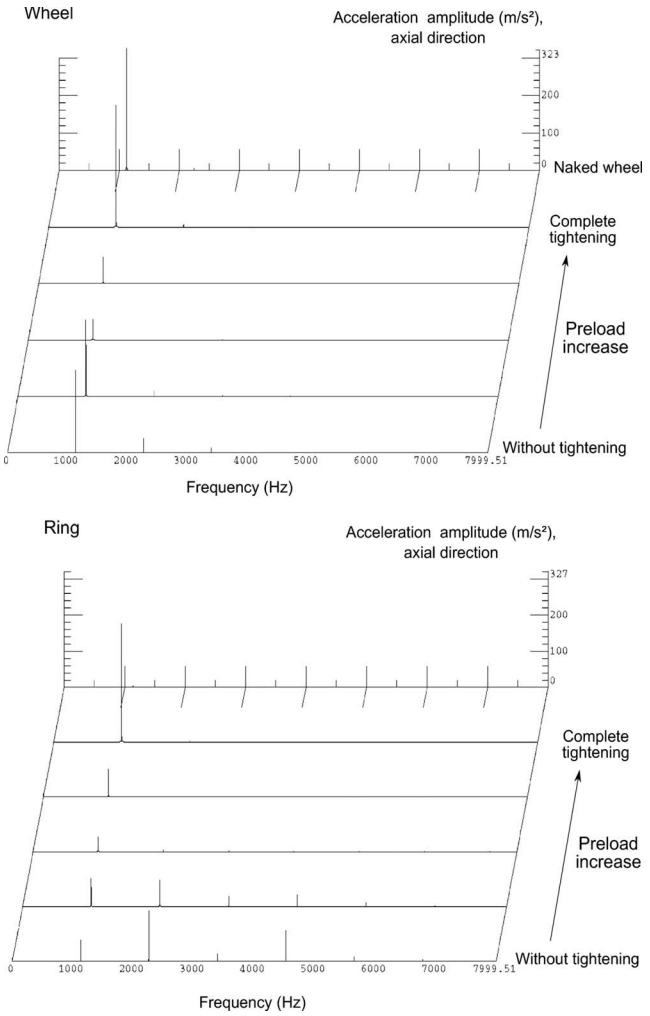


FIG. 7. Waterfall plots of wheel and ring acceleration for a sinusoidal excitation of the 0-3 mode.

ring can be seen. Note that harmonics are observed in the ring acceleration spectra for low preload. Figure 8 shows wheel and ring time acceleration for low and high preload. For low preload, ring and wheel vibratory responses are uncoupled, with independent dynamic motions contrary to the coupled motion obtained for high preload.

Attenuation mechanism seems to progress coupling between the ring and the wheel circumferentially with tightening increase. It may explain the presence of harmonics for low preload.

C. Summary of the experimental results

Several points have been highlighted during the experimental investigations:

- existence of a range of optimal preload values leading to the lowest sound pressures between the extreme configurations of none and complete tightening of the ring,
- a sound pressure level greater than the naked wheel (without ring) when the preload is maximum corresponding to a full tightening of the ring,
- modification of the wheel frequencies by varying preload, and

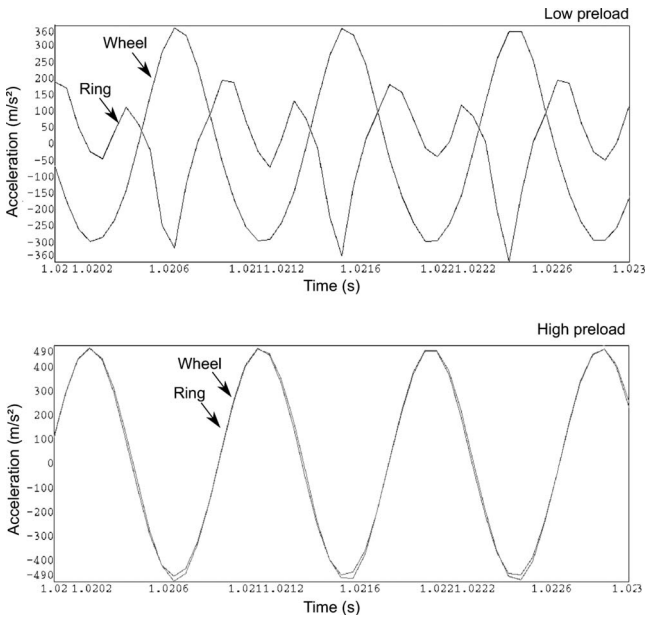


FIG. 8. Wheel and ring acceleration time history for a sinusoidal excitation of the 0-3 mode and for low and high preloads.

- attenuation of both the wheel and ring acceleration levels around the optimal preload values.

Note that none of these observations refer to impact or friction phenomena between the ring and the groove. It appears that as the preload increases the ring-wheel system changes from one in which the ring and wheel are uncoupled to one in which the two are fully coupled. To help explain these experimental results, we now turn to a simplified theoretical model in Sec. IV.

IV. THEORETICAL MODEL OF THE ATTENUATION MECHANISM

In this section, we simulate the wheel-ring system with a one-dimensional mass-spring model to illustrate the mechanism of attenuation. A comprehensive a 3D numerical model will be presented in Sec. V.

The mass-spring model is based on a two two-degree-of-freedom double mass-spring system, as shown in Fig. 9. The first mass M_w refers to the wheel and the second M_r refers to the ring. Wheel and ring displacements are noted X_w and X_r, respectively.

Mass, stiffness, and damping parameters are listed in Table I. They have been adjusted to obtain the same resonance frequency of the experimental wheel eigenmode: 0-3 at around 1100 Hz. Note that the damping ratio for the naked wheel is around 0.02%. Preload effects are introduced by the

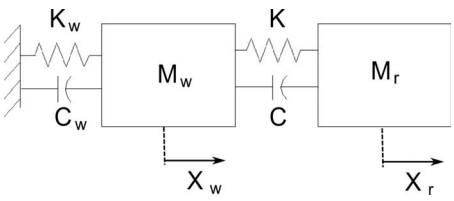


FIG. 9. 2 DOF model of the ring-damped wheel.

TABLE I. Parameters used in the mass-spring model.

M_w	300 kg
M_r	9 kg
K_w	1.5×10^{10} N/m
C_w, C	425 N s/m

stiffness between the ring and the wheel. The relationship between wheel-ring contact stiffness and preload is not known but is assumed that the stiffness increases monotonically with load. Finally, K and C correspond to contact stiffness and damping, respectively, of the ring and wheel assembly. Note that the natural frequency of the ring is not considered in the analytical approach. The main assumption is the contact induced stiffness is greater than the ring stiffness. The analytical model is only presented here to explicit the coupling mechanism but it is clear that it can be improved by introducing the mass-spring-damped system for the ring, which would eliminate the natural frequency of the ring going to zero when the contact spring stiffness goes to zero. However, it needs to characterize the ring stiffness.

Displacement amplitudes of the wheel and the ring are given by

$$\left|\frac{X_w}{F}\right| = \left|\frac{K + iC\omega - \omega^2 M_r}{(K_w + K - M_w\omega^2 + i(C_w + C)\omega)(K - M_r\omega^2 + iC\omega) - (K + iC\omega)^2}\right|,$$

$$\left|\frac{X_r}{F}\right| = \left|\frac{K + iC\omega}{(K_w + K - M_w\omega^2 + i(C_w + C)\omega)(K - M_r\omega^2 + iC\omega) - (K + iC\omega)^2}\right|. \quad (1)$$

Natural frequencies are given by the roots of the characteristic polynomial of the system (1), damping effects are neglected. Figure 10 shows the change in the natural frequencies with respect to the contact stiffness K.

For low values of K, the natural frequencies of the two systems (wheel and ring) are well separated. Wheel and ring natural frequencies increase with contact stiffness, with the ring increasing more rapidly.

For a contact stiffness value of $K=4.5\times10^8$ N/m, the main mass amplitude is zero. For higher stiffnesses, the natural frequencies are for the coupled wheel-ring system and the ring alone. As observed in the experimental investigations, the natural frequencies increase with increasing contact stiffness (increasing preload). Around the transition zone, a bifurcation of the wheel natural frequency to a wheel-ring system appears. Two close natural frequencies exist. This point is illustrated in Fig. 11 by the amplitude of vibration of each component. In the transition zone (around the white line in Fig. 11), both wheel and ring accelerations decrease.

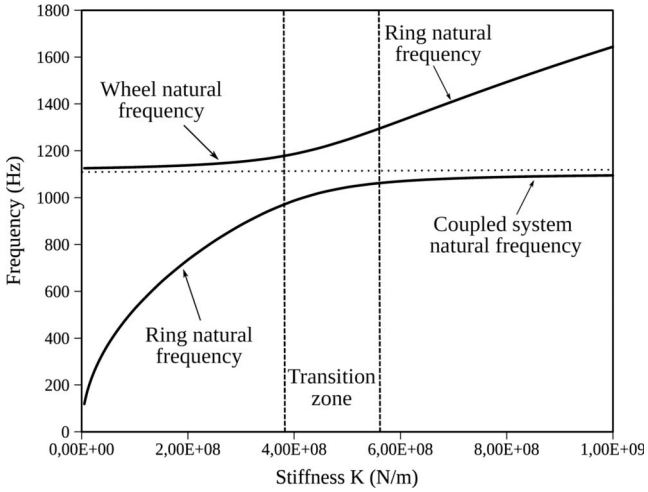


FIG. 10. Natural frequencies vs contact stiffness.

K corresponds to global value of the wheel-ring assembly considering that the ring stiffness is much lower the induced contact stiffness. However, the natural frequency of the ring is not described separately and goes to zero when the contact spring stiffness goes to zero (Figs. 10 and 11). This model may be improved by introducing a mass-spring-damped system for the ring providing its characterization.

Note that this simple model gives an explanation of most of the experimental observations: variations in the wheel natural frequency with preload, appearance of many peaks around the optimal value of preload and attenuation of acceleration levels of the wheel as well as those of the ring.

The simplified model helps to explain the origin of the attenuation of vibration as a stiffness modification leading to a bifurcation of the wheel mode to a wheel-ring mode. This bifurcation results in a decrease in the level of acceleration of both components.

V. 3D FEM APPLICATION

In this section a 3D FE model of the wheel equipped with an external ring, is used in a prestressed modal analysis to determine the natural frequencies of the system. The geometry of the model is presented in Fig. 12(a) and is described in (Brunel, 2004a, 2004b). Contact between the ring and the groove is modeled by surface-surface contact elements with a penalty algorithm (contact stiffness is introduced when the contact element status is closed allowing the ring-wheel contact forces in all directions to be included in the analysis). In this investigation the modal analysis is focused around the mode (1100 Hz) that was investigated experimentally. In addition, the normal stiffness of the contact

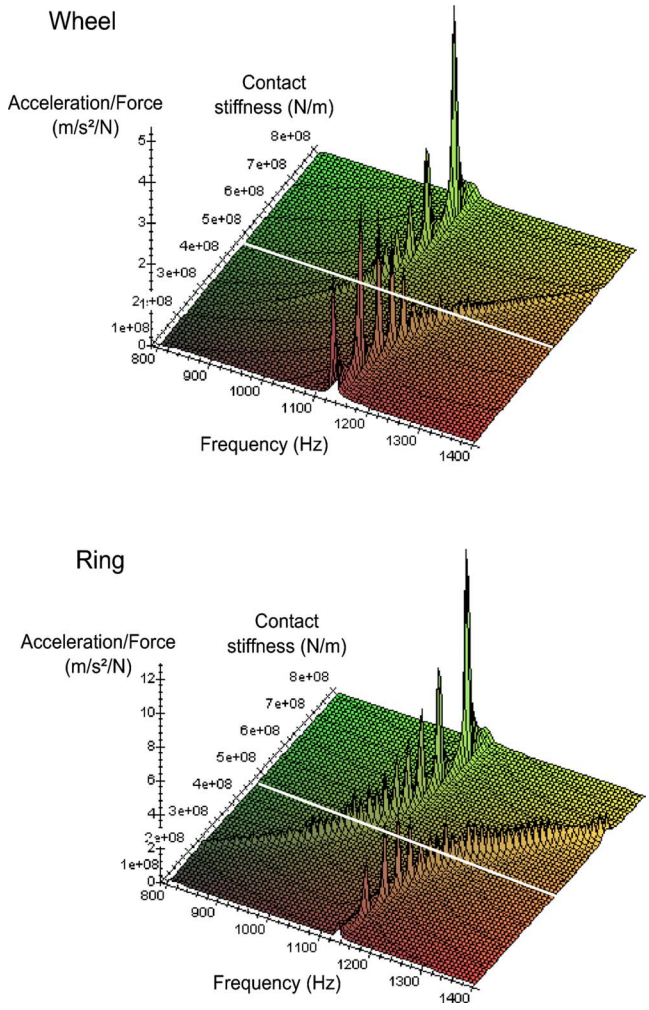


FIG. 11. (Color online) Wheel and ring acceleration response vs contact stiffness.

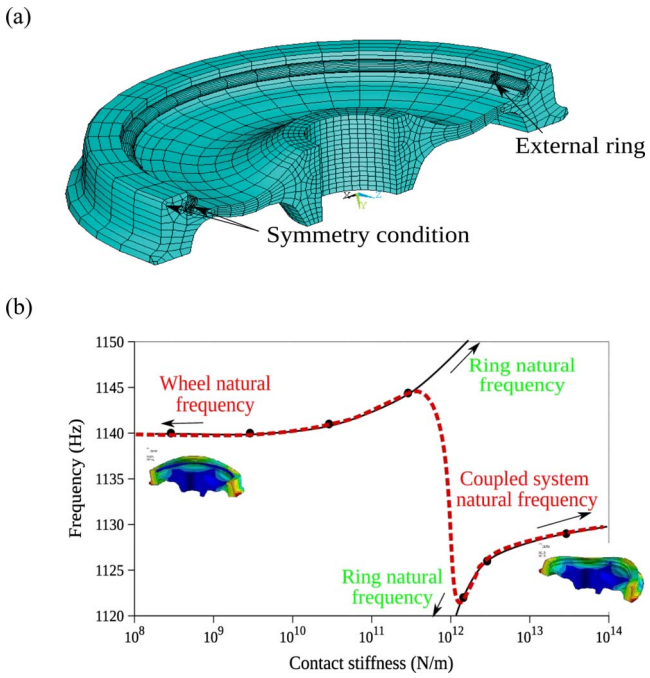


FIG. 12. (Color online) (a) 3D FE model of the ring-damped wheel. (b) 0–3 mode frequency vs contact stiffness.

elements has been varied between 1×10^9 and 1×10^{14} N/m, simulating the preload variation in the experiment.

Figure 12(b) shows that the same vibration behavior is obtained with this more complex model as was obtained with the previous simplified theoretical model. Increasing the contact stiffness causes the wheel natural frequency first to increases, followed by a decrease corresponding to a bifurcation to a coupled mode. Further increases in contact stiffness then cause the coupled wheel-ring frequency to increases. Pictures of the finite element model (FEM) results show that above the contact stiffness that results in bifurcation, the wheel and the ring vibrate together. Note that this numerical simulation also shows a jump in frequency of about 25 Hz (in the bifurcation region) that agrees with that observed experimentally.

VI. CONCLUSIONS

This paper presents a study of the effect of preloaded metal rings inserted into grooves in railway wheels. The aim is to design optimized railway wheels which are less noisy during squeal. A combined theoretical and experimental approach was taken. Experiments indicated that there is an "optimal preload," where the response of the wheel-ring system (measured in terms of axial accelerations and also in terms of radiated sound pressure) to white noise is smaller than that of the wheel without ring. A theoretical model of the wheel-ring system was created that consisted of a one-dimensional two-degree-of-freedom double mass-spring system with a spring of variable stiffness connecting the two masses. The resonance frequencies of this system were calculated, and it was found that for low stiffness the two masses move largely independently, while coupled motion was observed for higher stiffness. A FEM model was also developed that gave three-dimensional displacements and forces in the wheel system and lead to the same kind of results, with more accurate values compared to experiments. It is remarkable that the model proposed here gives the same conclusions as all of the experimental observations: modal frequency variation, optimal stiffness/preload (value corresponding to the lowest acceleration of both the wheel and the ring), and highest noise level if preload is maximum.

It is concluded that the transition from independent to coupled motion is the mechanism responsible for the reduced wheel noise at the optimal preload. It also confirms previous results of a secondary influence of friction dissipation at the wheel-ring contact.

The models can now be used to analyze the impact of the damping ring on each mode of the wheel. They also allow a sensitivity analysis to be carried out to determine the key parameters giving maximal attenuation.

Furthermore knowledge of the attenuation mechanism can be used to open new ways to improve the reliability and the efficiency of the ring dampers. For example, it explains that connecting the extremities of the ring by welding gives variable noise reduction performance because the preload is not controlled. For more reliable performance, the preload

obtained during the assembly process has to be controlled. Several ways are under investigation for improving the noise reduction performance of the damping rings, such as by using different ring geometries and by controlling values of the ring preload to attenuate the noisiest modes.

ACKNOWLEDGMENTS

The present research work has been supported by International Campus on Safety and Intermodality in Transportation the Nord-Pas-de-Calais Region, the European Community, the Regional Delegation for Research and Technology, the Ministry of Higher Education and Research, and the National Center for Scientific Research. The authors gratefully acknowledge the support of these institutions.

Akay, A. (**2002**). "Acoustics of friction," J. Acoust. Soc. Am. **111**, 1525–1548.

Brunel, J.-F. (**2004b**). "Etude vibro-acoustique du crissement des roues ferroviaires et de son atténuation par insertion d'éléments métalliques (Vibro-acoustic analysis of railway wheel squeal noise and attenuation by metallic rings)," Ph.D. thesis, Université des Sciences et Technologies de Lille, Lille, France.

Brunel, J.-F., and Dufrénoy, P. (**2007**). "Influence de la dynamique locale de contact sur la réponse vibratoire globale d'un systéme: Application au crissement (Consequences of self excited contact vibration to dynamic behavior of structures: Application to squeal noise)," in Eighth Colloque National en Calcul des Structures, Giens, France, pp. 375–381.

Brunel, J.-F., Dufrénoy, P., Charley, J., and Demilly, F. (**2003**). "Attenuation of the squeal noise of railway wheels using metallic rings," in Tenth International Congress on Sound and Vibration Proceedings, Stockholm, Sweden, pp. 1681–1688.

Brunel, J.-F., Dufrénoy, P., and Demilly, F. (**2004a**). "Modelling of squeal noise attenuation of ring damped wheels," Appl. Acoust. **65**, 457–471.

Brunel, J.-F., Dufrénoy, P., Nait, M., Munoz, J.-L., and Demilly, F. (**2006**). "Transient models for curve squeal noise," J. Sound Vib. **293**, 758–765.

López, I. (**1999**). "Theoretical and experimental analysis of ring-damped wheels," Ph.D. thesis, University of Navarra, Navarra, Spain.

López, I., Busturia, J.-M., and Nijmeijer, H. (**2004**). "Energy dissipation of a friction damper," J. Sound Vib. **278**, 539–561.

Rudd, M. J. (**1976**). "Wheel/rail noise, Part II: Wheel squeal," J. Sound Vib. **46**, 381–394.

Thompson, D. J., and Jones, C. J. C. (**2000**). "A review of the modelling of wheel-rail noise generation," J. Sound Vib. **231**, 519–536.

Vincent, N., Koch, J.-R., Chollet, H., and Guerder, J.-Y. (**2004**). "Curve squeal of urban rolling stock. Step 1: State of the art and field measurements," Eighth Workshop on Railway Noise Proceedings, Buxton, England, pp. 201–212.

Wetta, P., and Demilly, F. (**1985**). "Reduction of wheel squeal generated on curves or during braking," Fifth International Wheelset Congress Proceedings, Paris, France, pp. 17–22.

Propagation of elastic waves in a fluid-loaded anisotropic functionally graded waveguide: Application to ultrasound characterization

Cécile Baron and Salah Naili[a)]
Laboratoire de Mécanique Physique, Faculté des Sciences et Technologie, Université Paris 12–Val de Marne, 61 Avenue du Général de Gaulle, 94010 Créteil Cedex, France

(Received 27 June 2008; revised 19 August 2009; accepted 22 December 2009)

Non-destructive evaluation of heterogeneous materials is of major interest not only in industrial but also in biomedical fields. In this work, the studied structure is a three-layered one: A laterally heterogeneous anisotropic solid layer is sandwiched between two acoustic fluids. An original method is proposed to solve the wave equation in such a structure without using a multilayered model for the plate. This method is based on an analytical solution, the matricant, explicitly expressed under the Peano series expansion form. This approach is validated for the study of a fluid-loaded anisotropic and homogeneous plane waveguide with two different fluids on each side. Then, original results are given on the propagation of elastic waves in an asymmetrically fluid-loaded waveguide with laterally varying properties. This configuration notably corresponds to the axial transmission technique to the ultrasound characterization of cortical bone *in vivo*.
 [DOI: 10.1121/1.3292949]

PACS number(s): 43.40.Dx, 43.20.Mv [ADP] Pages: 1307–1317

I. INTRODUCTION

A lot of natural media have unidirectional varying elastic properties. The mantel crust, the oceans, and cortical bone are some of these functionally graded media. Scientists focused on the advantages presented by this type of materials in terms of mechanical behavior, and since the 1980s, they developed industrial functionally graded materials (FGMs) particularly exploited in high-technology and biomedical applications. Consequently, the non-destructive evaluation of these materials is a key issue. Surface and guided waves play a major role in non-destructive testing and evaluation of complex structures. Several studies are dedicated to the leaky Lamb wave propagation in fluid-loaded plates (Chimenti and Nayfeh, 1990; Chimenti and Rokhlin, 1990; Deschamps and Poncelet, 2000). In all these studies, the media are homogeneous or multilayered. In this work, we introduce a general method to take into account the continuity of the property variation in an anisotropic waveguide. This method is based on the knowledge of an analytical solution of the wave equation, the matricant, explicitly expressed via the Peano series (Peano, 1888). The accuracy of the numerical evaluation of this solution and its validity domain are perfectly managed (Baron, 2005; Youssef and El-Arabawi, 2007). Because it deals with an analytical solution, all the wave propagation parameters are controlled. This represents an advantage with respect to numerical methods such as finite-element and finite-difference methods for which the problem treated is a global one, making difficult to analyze and interpret the experimental data, which result from the interaction and coupling of numerous physical phenomena. To the best of our knowledge, this is the first method to evaluate the mechanical behavior of a fluid-loaded anisotropic waveguide with continuously varying properties without modeling the FGM plate by a multilayered plate. Consequently, in the context of real materials with continuously varying properties (such as bone, bamboo, or manufactured FGM), this method could allow assessing the solution to a more realistic model with a controlled accuracy and without an increase in the computation-time.

In this work, we first present the method and its setup with fluid-structure interaction; then we proceed to the validation of the method by comparing our results to the dispersion curves obtained from classical schemes on homogeneous waveguides (isotropic and anisotropic). Two advantages of the method are underlined: (i) An asymmetric fluid-loading may be taken into account without modifying the scheme for the numerical solution, and (ii) the influence of the property gradient on the ultrasonic response of the waveguide may be investigated via the frequency spectrum of the reflection coefficient modulus. Finally, we get onto the relevancy of this model applied to the ultrasound characterization of cortical bone by the axial transmission technique.

II. BACKGROUND

Contemporary work efforts over the last 2 decades illustrate some of the technology interest on guided waves to non-destructive evaluation. Namely, Rose (2002) gave a review of ultrasonic guided wave inspection potential. A lot of papers deal with the interaction between guided waves and a solid plate immersed in a fluid or embedded between two different fluids. Guided modes in an infinite elastic isotropic plate in vacuum were first treated by Rayleigh (1885) then by Lamb (1917). The Lamb wave problem is reserved, strictly speaking, for wave propagation in a traction-free homogeneous isotropic plate. To deal with guided modes in a fluid-loaded plate, we use the term "leaky Lamb waves" as the energy is partly radiated in the fluids on both sides of the plate. For the basic Lamb problem—plate in vacuum—the

[a)]Author to whom correspondence should be addressed. Electronic mail: naili@univ-paris12.fr

solutions of the dispersion equation are real whereas in the case of a plate bounded by media on both sides, the dispersion equation has complex solutions. In 1961, Worlton (1961) gave an experimental confirmation of the theoretical work of Lamb (1917), by obtaining experimentally the dispersion curves of aluminum and zirconium plates, asserting that water loading has little effect on the behavior of waves in plates. In 1976, Pitts *et al.* (1976) presented some numerical test results on the relationship between real part of the reflection coefficient poles and the phase velocity of leaky Lamb modes in a homogeneous isotropic brass plate in water. Folds and Loggins (1977) proposed analytical expressions of the reflection and transmission coefficients for plane waves at oblique incidence on a multilayered isotropic plate immersed in water based on Brekhovskikh's (1980) analysis. They found good agreement with their theoretical results and experimental data. Few years later, Fiorito *et al.* (1979) developed a resonance formalism for the fluid-loaded elastic plate and gave some theoretical and numerical results for an isotropic steel plate immersed in the water. This formalism was generalized to the interactions of acoustic plane waves with an asymmetrically fluid-loaded elastic plate by Franklin *et al.* (2001). Nayfeh, Chimenti, and Rokhlin produced a lot of works on wave propagation in anisotropic media and particularly in fiber composite plates immersed in a fluid (Chimenti and Nayfeh, 1986, 1990; Chimenti and Rokhlin, 1990; Nayfeh and Chimenti, 1988, 1989; Rokhlin and Wang, 2002). Based on their formalism, Deschamps and Poncelet (2000) placed the emphasis on the difference between what they called transient Lamb wave-solutions of the characteristic equation of the plate for complex frequency and real slowness (time attenuation), and heterogeneous Lamb waves for which the slowness is complex and the frequency is real (spatial attenuation). These two ways of resolution of the dispersion equation have two different physical meanings—space or time attenuation—and consequences of this difference are developed in their paper (Deschamps and Poncelet, 2000). A critical point is the validity of the Cremer's coincidence hypothesis: The real couples (angular frequency ω and phase velocity v_φ), such that the reflection coefficient is minimum, may be identified as velocity dispersion of plate waves. Experimentally checked in a lot of configurations, it appears to be not well satisfied in several cases (for instance, graphite-epoxy plates when the ratio between fluid and plate mass densities is not "small") (Chimenti and Nayfeh, 1986; Nayfeh and Chimenti, 1988). The results obtained by Deschamps and Poncelet (2000) on fluid-loaded plate show a good correlation between dispersion curves obtained in complex frequency and the minima of the reflection coefficient, which suggests that the Cremer's coincidence principle is still valid considering time attenuation. All these studies show the evidence that the wave propagation in fluid-loaded elastic plate emerges as a very delicate problem, which needs cautious treatment.

III. GENERAL FORMULATION OF THE PROBLEM

We consider a three-dimensional multilayer system composed of one elastic solid layer sandwiched between two acoustic fluid layers. Let $\mathbf{R}(\mathrm{O},\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$ be the Cartesian frame of reference where O is the origin of the space and $(\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3)$ is an orthonormal basis for this space. The coordinate of the generic point $\mathbf{x}$ in $\mathbf{R}$ is specified by (x_1,x_2,x_3). The acoustic fluid layers occupy an open unbounded domain. Both fluids f_1 and f_2 are supposed perfect, of respective mass densities ρ_{f_1} and ρ_{f_2}; the constant speeds of sound in each fluid are c_{f_1} and c_{f_2}, respectively. The thickness of the solid layer is denoted by d and its mass density by ρ. The interfaces between the fluids and the solid layer are infinite planes parallel to the $(\mathbf{x}_1,\mathbf{x}_2)$-plane. The $\mathbf{x}_3$-axis is oriented downward and the origin O is located at the interface between the upper fluid f_1 and the solid layer. Therefore, we assume that the structure is a two-dimensional one and that the guided waves travel in the plane $x_2=0$; in the following parts, this coordinate is implicit and is omitted in the mathematical expressions. Moreover, the solid layer will be so-called plate.

The elastic plate is supposed to be anisotropic and is liable to present continuously varying properties along its thickness ($\mathbf{x}_3$-axis). These mechanical properties are represented by the stiffness fourth-order tensor $\mathbb{C}=\mathbb{C}(x_3)$ and the mass density $\rho=\rho(x_3)$.

A. System equations

1. The wave equation in the fluid f_n (for n=1 or 2)

In the fluid f_n and the context of the linear acoustic theory, the linearized Euler's equation and the constitutive equations are written as

$$-\frac{\partial p^{(n)}}{\partial x_j}=\rho_{f_n}\frac{\partial^2 u_j^{(n)}}{\partial t^2},$$

$$p^{(n)}=K_{f_n}\,\mathrm{div}\,\mathbf{u}^{(\mathbf{n})}, \tag{1}$$

where $\mathbf{u}^{(n)}$ and $p^{(n)}$, respectively, represent the displacement vector and the pressure in the fluid f_n; its compressibility and the speed of sound in the fluid at equilibrium are, respectively, K_{f_n} and $c_{f_n}=\sqrt{K_{f_n}/\rho_{f_n}}$. The operator div is the divergence.

The solutions of system (1) for the fluid f_n are sought under the form

$$\mathbf{f}_n(x_1,x_2;t)=\mathbf{A}_n(x_3)\exp\,\imath(k_1x_1+k_3^{(n)}x_3-\omega t), \tag{2}$$

where k_1 and $k_3^{(n)}$ are the wavenumbers, respectively, along the $\mathbf{x}_1$-axis and $\mathbf{x}_3$-axis in the fluid f_n; the angular frequency is noted ω and $\imath$ is the imaginary unit.

We consider an incident wave reaching the plate at an angle θ_1 from the $\mathbf{x}_3$-axis in the fluid f_1. The incident displacement-field is defined in the following form, assuming that its amplitude is normalized:

$$\mathbf{u}_I^{(1)}=\begin{pmatrix}\sin\,\theta_1\\0\\\cos\,\theta_1\end{pmatrix}\exp\,\imath(k_1x_1+k_3^{(1)}x_3-\omega t), \tag{3}$$

with $k_1=(\omega/c_{f_1})\sin\,\theta_1$ and $k_3^{(1)}=(\omega/c_{f_1})\cos\,\theta_1$. From this, the expressions of the reflected displacement-field $\mathbf{u}_R^{(1)}$ in f_1 and of the transmitted displacement-field $\mathbf{u}_T^{(2)}$ in f_2 are deduced as follows:

$$\mathbf{u}_R^{(1)} = R\begin{pmatrix} \sin\theta_1 \\ 0 \\ -\cos\theta_1 \end{pmatrix} \exp \iota(k_1 x_1 - k_3^{(1)} x_3 - \omega t),$$

$$\mathbf{u}_T^{(2)} = T^{\frac{c_{f_2}}{c_{f_1}}}\begin{pmatrix} \sin\theta_1 \\ 0 \\ \cos\theta_1 \end{pmatrix} \exp \iota(k_1 x_1 + k_3^{(2)} x_3 - \omega t). \quad (4)$$

where $R=R(x_1,x_3;t)$ and $T=T(x_1,x_3;t)$, respectively, represent the reflection and transmission coefficients, which will be expressed explicitly in the sequel. The incident, reflected, and transmitted pressure fields, respectively, noted $p_I^{(1)}$, $p_R^{(1)}$, and $p_T^{(2)}$, are deduced from expressions (3) and (4), and the second equation of the system (1)

$$p_I^{(1)} = -\iota\omega \times \rho_{f_1} c_{f_1} \times \exp \iota(k_1 x_1 - k_3^{(1)} x_3 - \omega t),$$

$$p_R^{(1)} = -\iota\omega \times \rho_{f_1} c_{f_1} \times R \times \exp \iota(k_1 x_1 - k_3^{(1)} x_3 - \omega t),$$

$$p_T^{(2)} = -\iota\omega \times \rho_{f_2} c_{f_2} \times T \times \exp \iota(k_1 x_1 + k_3^{(2)} x_3 - \omega t). \quad (5)$$

B. The wave equation in the plate waveguide

The body forces in the solid plate are neglected. The balance equation of linear momentum associated with the constitutive law of linear elasticity (Hooke's law) gives the following equations:

$$\frac{\partial \sigma_{ij}}{\partial x_j} = \rho \frac{\partial^2 u_i}{\partial t^2},$$

$$\sigma_{ij} = \frac{1}{2} C_{ijk\ell}\left(\frac{\partial u_k}{\partial x_\ell} + \frac{\partial u_\ell}{\partial x_k}\right), \quad (6)$$

where u_i (for $i=1,\ldots,3$) and σ_{ij} (for $i,j=1,\ldots,3$), respectively, represent the components of the displacement-field $\mathbf{u}$ and of the stress $\boldsymbol{\sigma}$. In system (6), Einstein's convention of summation on repeated indices is used. The solutions are sought for the vectors of displacement $\mathbf{u}$ and traction σ_{i3} (for $i=1,\ldots,3$) (assumed to be harmonic in time t and space along the $\mathbf{x}_1$-axis) under the form

$$\mathbf{f}(x_1,x_3;t) = \mathbf{A}(x_3)\exp \iota(k_1 x_1 - \omega t). \quad (7)$$

C. Fluid-loading interface conditions

The conditions at both interfaces $x_3=0$ and $x_3=d$ are the continuity of the normal displacement and the continuity of the normal stress. We consider that the fluids f_1 and f_2 are perfect; consequently, the shear stresses are zero at the interfaces $[\sigma_{13}(x_1,0;t)=\sigma_{13}(x_1,d;t)=0$ and $\sigma_{23}(x_1,0;t)=\sigma_{23}(x_1,d;t)=0]$. The following relations are obtained:

$$u_3(x_1,0;t) = u_3^{(1)}(x_1,0;t), \quad u_3(x_1,d;t) = u_3^{(2)}(x_1,d;t),$$

$$\sigma_{33}(x_1,0;t) = -p^{(1)}(x_1,0;t), \quad \sigma_{33}(x_1,d;t) = -p^{(2)}(x_1,d;t), \quad (8)$$

with

$$u_3^{(1)} = \mathbf{u}_I^{(1)} \cdot \mathbf{x}_3 + \mathbf{u}_R^{(1)} \cdot \mathbf{x}_3, \quad u_3^{(2)} = \mathbf{u}_T^{(2)} \cdot \mathbf{x}_3 \quad \text{and}$$

$$p^{(1)} = p_I^{(1)} + p_R^{(1)}, \quad p^{(2)} = p_T^{(1)}. \quad (9)$$

D. A closed-form solution: The matricant

Introducing expression (7) in Eq. (6), we obtain the wave equation under the form of a second-order differential equation with non-constant coefficients. For particular forms of profiles, this equation has analytical solutions expressed with special functions (Bessel or Hankel functions) (Vlasie and Rousseau, 2004). But, in the general case, there is no analytical solution to the problem thus formulated. The most current methods to solve the wave equation in unidirectionally heterogeneous media are derived from the Thomson–Haskell method (Haskell, 1953; Thomson, 1950). These methods are appropriate for multilayered media (Hosten and Castaings, 2003; Kenneth, 1982; Lévesque and Piché, 1992; Wang and Rokhlin, 2001). But, for continuously varying media, these techniques mean to replace the continuous profiles of properties with step-wise functions. Thereby, the studied problem becomes an approximate one, even before the resolution step; the accuracy of the solution as its validity domain is hard to evaluate. Moreover, the multilayered model of the waveguide creates some "virtual" interfaces likely to induce artifacts. In order to deal with the exact problem, that is to keep the continuity of the properties' variation, the wave equation is re-written under the form of an ordinary differential equation system with non-constant coefficients for which an analytical solution exists: the matricant (Baron, 2005).

1. Hamiltonian form of the wave equation

We consider that the plate presents material symmetries that allow decoupling the pressure–shear vertical (P-SV) waves, polarized in the propagation plane $(\mathbf{x}_1,\mathbf{x}_3)$ and the shear horizontal (SH) waves polarized along $\mathbf{x}_2$-axis. The incident medium f_1 is a perfect fluid; only the P-SV waves travel in the plate. Applying a spatio-temporal Fourier transform on (x_1,t) of the displacement-field [noted $\hat{\mathbf{u}}(k_1,x_3;\omega)$ after Fourier transform] and on the traction field [noted $\hat{\sigma}_{i3}(k_1,x_3;\omega)$ for $i=1,\ldots,3$] and using the Voigt notation ($C_{ijk\ell}$ for $i,j,k,\ell=1,\ldots,3$ is replaced with c_{IJ} for $I,J=1,\ldots,6$), Eq. (6) leads to Stroh's (1962) sextic plate formalism (Hamiltonian formulation of the wave equation)

$$\frac{\partial \hat{\sigma}_{13}}{\partial x_3} = \rho(\iota\omega)^2 \hat{u}_1 - \iota k_1 \hat{\sigma}_{11}, \quad \frac{\partial \hat{\sigma}_{33}}{\partial x_3} = \rho(\iota\omega)^2 \hat{u}_3 - \iota k_1 \hat{\sigma}_{13}, \quad (10)$$

$$\hat{\sigma}_{11} = \iota k_1 c_{11} \hat{u}_1 + c_{33}\frac{\partial \hat{u}_3}{\partial x_3}, \quad \hat{\sigma}_{13} = c_{55}\left(\frac{\partial \hat{u}_1}{\partial x_3} + \iota k_1 \hat{u}_3\right),$$

$$\hat{\sigma}_{33} = \iota k_1 c_{13} \hat{u}_1 + c_{33}\frac{\partial \hat{u}_3}{\partial x_3}. \quad (11)$$

According to Eqs. (10) and (11), $\hat{\sigma}_{11}$ is function of $\hat{u}_1$ and $\hat{\sigma}_{33}$. The wave equation becomes a matrix system expressed using the Thomson–Haskell parametrization of Stroh's (1962) sextic plate formalism

$$\frac{d}{dx_3}\eta(x_3) = \iota\omega\mathbf{Q}(x_3)\,\eta(x_3), \tag{12}$$

that is,

$$\frac{d}{dx_3}\begin{pmatrix} \iota\omega\hat{u}_1 \\ \iota\omega\hat{u}_3 \\ \hat{\sigma}_{13} \\ \hat{\sigma}_{33} \end{pmatrix} = \iota\omega\begin{pmatrix} 0 & s_1 & 1/c_{55}(x_3) & 0 \\ -s_1c_{13}(x_3)/c_{33}(x_3) & 0 & 0 & 1/c_{33}(x_3) \\ \rho(x_3)-s_1^2\zeta(x_3) & 0 & 0 & -s_1c_{13}(x_3)/c_{33}(x_3) \\ 0 & \rho(x_3) & -s_1 & 0 \end{pmatrix}\begin{pmatrix} \iota\omega\hat{u}_1 \\ \iota\omega\hat{u}_3 \\ \hat{\sigma}_{13} \\ \hat{\sigma}_{33} \end{pmatrix}, \tag{13}$$

with the relations

$$\zeta(x_3) = c_{11}(x_3) - \frac{c_{13}^2(x_3)}{c_{33}(x_3)}, \quad k_1 = \omega s_1, \tag{14}$$

where s_1 is the x_1-component of the slowness. The matrix $\mathbf{Q}$ includes all the information about the heterogeneity of the waveguide because it is expressed from the plate mechanical properties $(\rho(x_3), \mathbb{C}(x_3))$ and from two acoustical parameters (s_1, ω).

2. Explicit solution: The Peano expansion of the matricant

The wave equation thus formulated has an analytical solution expressed between a reference point $(x_1, 0, x_3^0)$ and some point of the plate $(x_1, 0, x_3)$ in the propagation plane. This solution is called the matricant and is explicitly written under the form of the Peano series expansion (Gantmacher, 1959; Peano, 1888; Pease, 1965):

$$\mathbf{M}(x_3, x_3^0) = \mathbf{I} + (\iota\omega)\int_{x_3^0}^{x_3}\mathbf{Q}(\xi)d\xi + (\iota\omega)^2 \times \int_{x_3^0}^{x_3}\mathbf{Q}(\xi)\left(\int_{x_3^0}^{\xi}\mathbf{Q}(\xi_1)d\xi_1\right)d\xi + \ldots, \tag{15}$$

where $\mathbf{I}$ is the identity matrix of dimension (4,4). If the matrix components $\mathbf{Q}(x_3)$ are bounded in the study interval, these series are always convergent (Baron, 2005). The components of the matrix $\mathbf{Q}$ are continuous in x_3 and the study interval is bounded (thickness of the waveguide); consequently the hypothesis is always verified. We underline that the $\iota\omega$-factorization leads up to a polynomial form of the matricant. The $\iota\omega$-polynomial coefficients are matrices independent of ω.

3. Boundary conditions: Fluid-structure interaction

Using the propagator property of the matricant through the plate thickness, the state-vector [defined in Eq. (13)] at the second interface $\eta(d)$ is evaluated from the state-vector at the first interface $\eta(0)$ as follows:

$$\eta(d) = \mathbf{M}(d,0)\,\eta(0). \tag{16}$$

The fluid-structure interaction sets the conditions of zero shear stresses (see Sec. III C), used after a spatio-temporal Fourier transform on (x_1, t). Equation (16) becomes

$$\begin{pmatrix} \iota\omega\hat{u}_1(k_1,d;\omega) \\ \iota\omega\hat{u}_3(k_1,d;\omega) \\ 0 \\ \hat{\sigma}_{33}(k_1,d;\omega) \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{12} & M_{22} & M_{23} & M_{24} \\ M_{13} & M_{23} & M_{33} & M_{34} \\ M_{14} & M_{24} & M_{34} & M_{44} \end{pmatrix} \times \begin{pmatrix} \iota\omega\hat{u}_1(k_1,0;\omega) \\ \iota\omega\hat{u}_3(k_1,0;\omega) \\ 0 \\ \hat{\sigma}_{33}(k_1,0;\omega) \end{pmatrix}. \tag{17}$$

The condition to obtain a nontrivial solution to Eq. (17) leads to the following relation:

$$\iota\omega\hat{u}_1(k_1,0;\omega)\times M_{13} + \iota\omega\hat{u}_3(k_1,0;\omega)\times M_{32} + \hat{\sigma}_{33}(k_1,0;\omega)\times M_{34} = 0, \tag{18}$$

where M_{ij} (for $i,j = 1,\ldots,4$) represent the components of the matrix $\mathbf{M}$. The displacement component $\hat{u}_1(k_1,0;\omega)$ can be expressed as a linear combination of $\hat{u}_3(k_1,0;\omega)$ and $\hat{\sigma}_{33}(k_1,0;\omega)$; thus system (16) of dimension 4 is reduced to a matrix system of dimension 2

$$\eta(d) = \begin{pmatrix} P_1 & P_2 \\ P_3 & P_4 \end{pmatrix}\eta(0) \quad \text{where} \quad \eta(x_3) = \begin{pmatrix} \iota\omega\hat{u}_3 \\ \hat{\sigma}_{33} \end{pmatrix}, \tag{19}$$

with the relations

$$P_1 = M_{22} - M_{21}\frac{M_{32}}{M_{31}}, \quad P_2 = M_{24} - M_{21}\frac{M_{34}}{M_{31}},$$

$$P_3 = M_{42} - M_{41}\frac{M_{32}}{M_{31}}, \quad P_4 = M_{44} - M_{11}\frac{M_{34}}{M_{31}}. \tag{20}$$

Interface conditions (8) are transformed in the Fourier domain (k_1, ω). The expressions of the displacement and the pressure in the fluids [see Eqs. (3)–(5)], so that the one of the displacement and traction fields in the solid plate [see Eq. (19)], are substituted in the transformed interface conditions.

Setting $\eta(0)=(\alpha_1,\alpha_2)^T \exp \imath(k_1 x_1-\omega t)$, where the superscript ^{T} designates the transpose operator, we obtain the following matrix equation:

$$\begin{pmatrix} \imath\omega s_3^{(1)} c_{f_1} & 1 & 0 & 0 \\ -\imath\omega\rho_{f_1} c_{f_1} & 0 & 1 & 0 \\ 0 & P_1 & P_2 & -\imath\omega s_3^{(2)} c_{f_2} \exp(\imath\omega s_3^{(2)} d) \\ 0 & P_3 & P_4 & -\imath\omega\rho_{f_2} c_{f_2} \exp(\imath\omega s_3^{(2)} d) \end{pmatrix} \begin{pmatrix} \hat{R} \\ \alpha_1 \\ \alpha_2 \\ \hat{T} \end{pmatrix} = \begin{pmatrix} \imath\omega s_3^{(1)} c_{f_1} \\ \imath\omega\rho_{f_1} c_{f_1} \\ 0 \\ 0 \end{pmatrix}, \tag{21}$$

where $\mathbf{s}^{(n)}=\mathbf{k}^{(n)}/\omega$ is the slowness-vector in the fluid f_n ($n=1$ or 2). The quantities $\hat{R}$ and $\hat{T}$ are, respectively, the reflection and transmission coefficients expressed in the Fourier domain: $\hat{R}=\hat{R}(k_1,x_3;\omega)$ and $\hat{T}=\hat{T}(k_1,x_3;\omega)$. The two first lines of system (21) express the boundary conditions at the first interface ($x_3=0$) and the two last lines those at the second interface ($x_3=d$) introducing the Fourier transform of expressions (3)–(5) in the following relations:

$$\eta(0)-\begin{pmatrix} \imath\omega\hat{u}_{3_R}^{(1)} \\ -\hat{p}_R \end{pmatrix}_{x_3=0} = \begin{pmatrix} \imath\omega\hat{u}_{3_I}^{(1)} \\ -\hat{p}_I \end{pmatrix}_{x_3=0},$$

$$\begin{pmatrix} P_1 & P_2 \\ P_3 & P_4 \end{pmatrix} \eta(0)-\begin{pmatrix} \imath\omega\hat{u}_{3_T}^{(2)} \\ -\hat{p}_T \end{pmatrix}_{x_3=d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{22}$$

where $\hat{u}_{3_R}^{(1)}$, $\hat{u}_{3_I}^{(1)}$, and $\hat{u}_{3_T}^{(2)}$ are the components along $\mathbf{x}_3$-axis of $\hat{\mathbf{u}}_R^{(1)}$, $\hat{\mathbf{u}}_I^{(1)}$, and $\hat{\mathbf{u}}_T^{(2)}$ vectors, respectively. Note the equality between the quantities $u_{3_T}^{(2)}$ and $u_3^{(2)}$ where this last is defined in Eq. (9).

4. Expression of the reflection and transmission coefficients

From Eq. (21), we deduce the analytical expressions of the reflection and transmission complex coefficients

$$\hat{R}(s_1,x_3;\omega)=\frac{(P_3-P_1Z_2+P_4Z_1-P_2Z_1Z_2)}{(P_3-P_1Z_2-P_4Z_1+P_2Z_1Z_2)},$$

$$\hat{T}(s_1,x_3;\omega)=-\frac{2Z_2(\rho_{f_1}c_{f_1}/\rho_{f_2}c_{f_2})(P_1P_4-P_2P_3)}{(P_3-P_1Z_2-P_4Z_1+P_2Z_1Z_2)} \times\exp(-\imath\omega s_3^{(2)}d), \tag{23}$$

with $Z_n=\rho_{f_n}/\sqrt{1/c_{f_n}^2-s_1^2}$ (for $n=1$ or 2) and $k_1=\omega s_1$.

IV. VALIDATION OF THE METHOD

The aim of this section is to check that the Peano expansion of the matricant is well-adapted to study fluid-loaded waveguides. We take into account the fluid-structure interaction in different configurations of homogeneous plates comparing the results obtained from the numerical implementation of the Peano expansion of the matricant to results taken from literature.

TABLE I. Elastic properties of transversely isotropic plate (with $c_{23}=c_{22}-2c_{44}$).

ρ (g cm^{-3})	c_{11} (GPa)	$c_{22}=c_{33}$ (GPa)	$c_{12}=c_{13}$ (GPa)	c_{44} (GPa)	$c_{66}=c_{55}$ (GPa)
1.85	23.05	15.1	8.7	3.25	4.7

The numerical evaluation of P_1, P_2, P_3, and P_4 requires us to truncate the Peano series and to numerically calculate the integrals. Thus, the error can be estimated and controlled (Baron, 2005). We retained 70 terms in the series and evaluate the integrals over 100 points using Simpson's rule (fourth-order integration method). These choices ensure the convergence of the solution and the accuracy of the results for a reasonable computation-time (never exceeding few minutes on a desktop computer). Expressions (23) give the frequency spectrum (modulus and phase) of the reflection coefficient for different incidences (s_1 varies from zero-normal incidence to $1/c_{f_1}$ corresponding to the critical incidence in the fluid f_1).

A lot of works detailed the relationship between the poles and the zeros of the reflection coefficient and the leaky Lamb wave dispersion curves (Chimenti and Rokhlin, 1990; Deschamps and Poncelet, 2000). The results of Sec. IV A compare the dispersion curves obtained by seeking the poles of reflection coefficient (23) and the results taken from literature or from closed-form solution.

A. Validation for a homogeneous and isotropic or anisotropic fluid-loaded plate

The method is tested by plotting the dispersion curves (variation of the phase velocity versus frequency-thickness product) for an isotropic aluminum plate immersed in water. The data in the paper of Chimenti and Rokhlin (1990) are used. The results obtained (not shown) by the present method are in perfect agreement with the results presented by them (Chimenti and Rokhlin, 1990).

As mentioned by Chimenti and Rokhlin (1990), there are few differences between the zeros' loci and the poles' loci for a plate immersed in a fluid whose the mass density is lower than the plate mass density. As underlined by several authors, fluid-load does have just a weak influence on guided wave traveling in the plate immersed in water.

Taking into account the anisotropy does not change the scheme for the numerical solution of wave equation with the matricant. We consider a transverse isotropic plate immersed in water ($\rho_f=1$ g cm^{-3}, $c_f=1.485$ mm μs^{-1}) whose properties are reported in Table I. For that configuration, Nayfeh and Chimenti (1989) developed a method to obtain an analytical expression of the reflection coefficient. By using the data from this paper, the results obtained (see Fig. 1) with the present method are in perfect agreement with theirs. The curves presented in Figs. 1(a) and 1(b) are superimposed and need to be presented separately.

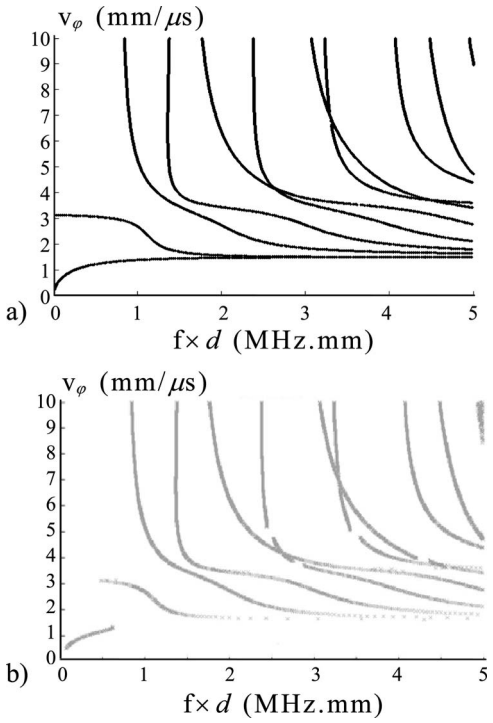


FIG. 1. Dispersion curves for a transversely isotropic plate immersed in water; comparison between (a) analytical results from Nayfeh and Chimenti (1989) and (b) results obtained via the Peano series expansion of the matricant.

B. Validation for an asymmetrically loaded homogeneous isotropic plate ($f_1 \neq f_2$)

The formalism presented here to solve the wave equation in an unidirectionally graded medium presents two main advantages: Without changing the scheme to obtain the numerical solution we can take into account (i) an asymmetric loading and (ii) the unidirectional continuous heterogeneity.

The mechanical behavior of the plate is different for symmetric and asymmetric loadings. For example, in the symmetric loading case, there is a unique critical frequency and a unique phase velocity value v_φ in the plate, which corresponds to the propagation velocity in the fluid ($v_\varphi = c_{f_1} = c_{f_2}$), for which the displacements and the stresses at the interfaces are quasi-null; whereas in the asymmetric loading ($f_1 = f_2$), there are two critical frequencies and two values of the phase velocity in the plate for which the structure does not respond (Dickey *et al.*, 1995). The validation is done on an isotropic aluminum plate with the following properties: $\rho = 2.79$ g cm^{-3}; the longitudinal and transverse wave velocities are, respectively, $v_L = 6.38$ mm μs^{-1} and $v_T = 3.10$ mm μs^{-1}. The characteristic properties of the fluid f_1 correspond to those of water (see Sec. IV A); the characteristic properties of the fluid f_2 correspond to glycerine: $\rho_{f_2} = 1.26$ g cm^{-3} and $c_{f_2} = 1.920$ mm μs^{-1}. This configuration is the same as the one studied by Franklin *et al.* (2001). The modulus of the reflection coefficient versus the incident angle is plotted in Fig. 2 for a fixed frequency-thickness product ($f \times d = 4.7$ MHz mm).

This figure shows the perfect agreement between our results and the ones presented by Franklin *et al.* (2001).

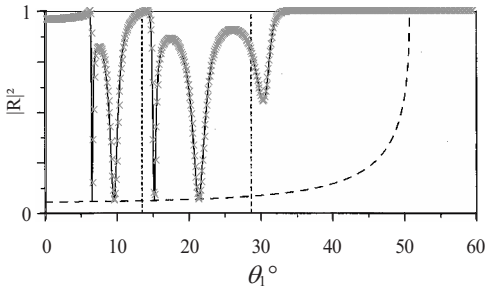


FIG. 2. Reflection coefficient modulus for an asymmetrically fluid-loaded aluminum plate (water and glycerine) versus incident angle: in dark lines results published in Franklin *et al.* (2001), and in gray crosses the reflection coefficient modulus calculated from Eq. (23). The vertical lines (dashed) represent the two critical angles for longitudinal waves and transverse waves. The dashed curve corresponds to the resonant amplitudes (Franklin *et al.*, 2001).

V. RELEVANCY OF THE METHOD FOR ULTRASOUND CHARACTERIZATION OF CORTICAL BONE

Cortical bone is a kind of hard tissue found at the edges of long bones and supports most of the load of the body. Several studies demonstrated the heterogeneous nature of the cortical bone, particularly they show evidence the gradual variation of the volumetric porosity (ratio between pores and total volume) along the cortical thickness. Yet, the porosity is intrinsically linked to the macroscopic mechanical behavior of the cortical bone (Baron *et al.*, 2007). Therefore, the continuous variation of porosity induces a continuous variation of material properties. Taking into account the gradient should prove itself to be essential in the context of diagnosis and therapeutic monitoring of osteoporosis. Indeed, the gradient characterization would allow assessing geometrical (cortex thickness) and material (elastic coefficients variation) information, which are fundamental parameters to evaluate the bone fragility. For several years, the quantitative ultrasonography (by axial and transverse transmissions) proved itself to be a hopeful alternative technique to evaluate the fracture risk (Marin *et al.*, 2006). However, the inter-individual and inter-site variations of bone mechanical properties make the standardization of the protocol of fracture risk evaluation by ultrasound very delicate.

The focus is set on a configuration closed to the axial transmission device for *in vivo* conditions. In this context, the relevancy of studying the head wave propagation has been demonstrated (Bossy *et al.*, 2004a, 2004b; Camus *et al.*, 2000). As a consequence all the reflection coefficients presented in this paper were calculated for an incident angle corresponding to the grazing-angle for longitudinal waves [critical angle of longitudinal wave propagation in the plate at the first interface ($x_3 = 0$)].

The surrounding media in the *in vivo* configuration of ultrasound characterization of cortical bone are the muscle for the upper fluid f_1 ($c_{f_1} = 1.54$ mm μs^{-1} and $\rho_{f_1} = 1.07$ g cm^{-3}) and the marrow for the lower fluid f_2 ($c_{f_2} = 1.45$ mm μs^{-1} and $\rho_{f_2} = 0.9$ g cm^{-3}) (Burlew *et al.*, 1980; Hill *et al.*, 1986). We are interested in the influence of the continuous gradient of the mechanical properties on the ultrasonic response in the configuration of *in vivo* cortical bone characterization.

A. Determination of a realistic range of variation of elastic bone properties

In order to define numerical values for a realistic value of the gradient of the different material properties, it is necessary to determine the limiting values reached by each elastic property. Our approach consists in considering *in vitro* measurements published in Dong and Guo, 2004 and performed in 18 samples. It is assumed that these limiting values for elastic properties are relevant for physiologic ranges of variations.

We assume that cortical bone is transversely isotropic. Transverse isotropy has been shown experimentally by different authors (Dong and Guo, 2004; Reilly and Burnstein, 1974; Rho, 1996) to be a realistic approximation of cortical bone degree of anisotropy.

Dong and Guo (2004) measured the homogenized bone properties by performing tensile and torsional tests with a mechanical testing system on 18 different human femoral bone specimens. The authors measured the values of the longitudinal and transverse Young's moduli (E_L and E_T, respectively) as well as the values of the longitudinal shear modulus G_L. From these measurements and by assuming constant values of Poisson's ratio, the values of the different components of the stiffness-tensor corresponding to the values of E_L, E_T, and G_L measured in Dong and Guo, 2004 were obtained following the relationships given in the Appendix.

The value of the longitudinal Poisson's ratio ν_L is taken equal to 0.37 for all computations because it corresponds to the average value found in Dong and Guo, 2004. The value of the transverse Poisson's ratio ν_T is taken equal to 0.45 following Eq. (A3) of the Appendix. The values of the stiffness coefficients corresponding to the mean values of the bone mechanical properties are referred to as "reference" set of parameters in what follows. The maximum and minimum values of the stiffness coefficients are obtained by considering, respectively, the maximum and minimum values of E_L and E_T within the range of variation measured in Dong and Guo, 2004, which is a simple way to obtain a realistic range of variation for the stiffness coefficients in cortical bone. Furthermore, the elastic properties deduced from the approach reported above were constrained to verify the thermodynamical stability conditions given in the Appendix by Eq. (A4).

We choose reference value of mass density ρ equal to 1.722 g cm^{-3}, following the value taken in Macocco *et al.*, 2006. In order to derive a realistic range of variation for mass density, we assume that the reference value is given by a porosity of 7%, which corresponds approximately to the mean porosity at the radius (Baron *et al.*, 2007). The porosity was assumed to vary between 3% and 15% (Bousson *et al.*, 2001; Dong and Guo, 2004) and a rule of mixture leads to the range of variation of mass density.

B. Modeling a gradient of material property

The impact of a controlled gradient vector $\boldsymbol{\delta}$ of any investigated material property S on the response of the structure studied is assessed. The scalar S corresponds to one of the stiffness coefficients c_{ij} of $\mathbb{C}$ or to mass density ρ. In each set of simulations, all the material properties are constant and equal to their reference value while S is subjected to the defined gradient.

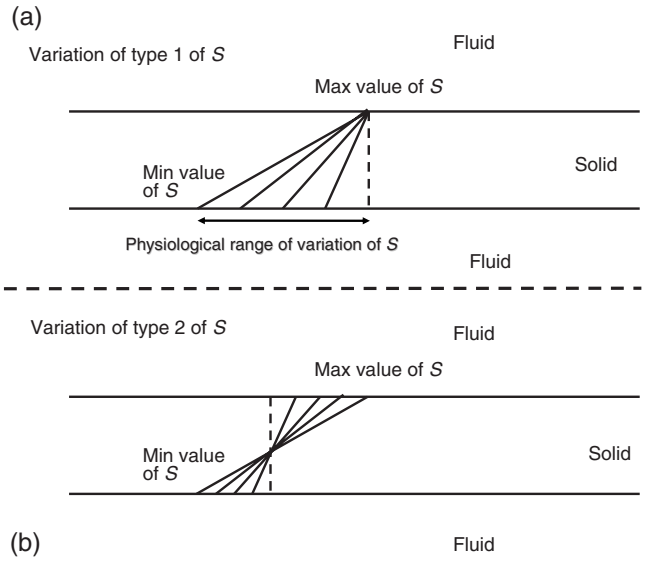


FIG. 3. Schematic representation of the two types of spatial variation considered for the material property S corresponding to the stiffness coefficient and to mass density. The solid lines in the solid layer indicate the spatial dependence of S. The dotted line in the solid layer indicates homogeneous material properties corresponding to the reference material properties. The variation of type 1 shown in (a) corresponds to a constant value at the bone-soft tissues' interface. The variation of type 2 shown in (b) corresponds to a constant value in the middle of the bone.

The gradient vector $\boldsymbol{\delta}=\text{grad}\, S=\delta \mathbf{x}_3$ is assumed to be independent of x_1 in all cases, where $\mathbf{x}_3$ is a unit vector along $\mathbf{x}_3$-axis and grad is the gradient operator acting on a scalar field. The quantity δ is always taken positive because the porosity is known to be higher in the endosteal part ($x_3=d$) than in the periosteal part ($x_3=0$) of the bone. Moreover, only the simple situation of affine spatial variations of S is considered, corresponding to a constant value of δ. This affine spatial variation of S is chosen because the actual physiological spatial dependence of S remains unknown. Two different affine spatial dependencies of the studied material property are considered and are illustrated in Fig. 3. Associated gradient δ will be referred to as type 1 or 2.

Type 1. The gradient of type 1 is such that the physical property S takes the same value S_M at the upper interface $x_3=0$ of the solid plate for all values of the gradient δ. The quantity $S(x_3)$ is therefore given by

$$S(x_3) = S_M + \delta \times x_3, \tag{24}$$

where S_M is given by the maximal value of the material property S considered. The maximal value δ_M of δ is chosen so that $S(d)$ is equal to S_m, where S_m is given by the minimal value of S. The gradient δ_M is given by

$$\delta_M = \frac{(S_m - S_M)}{d}. \tag{25}$$

Type 2. The gradient of type 2 is such that the physical property S takes the same value at the middle $x_3=d/2$ of the solid plate for all values of gradient δ. Furthermore, the

TABLE II. Mean value, maximum, and minimum values of the homogenized longitudinal and transversal Young moduli, of the four elastic constants and of mass density affecting the ultrasonic propagation in the framework of the model. These values are taken from Dong and Guo (2004).

Mechanical quantity	E_L (GPa)	E_T (GPa)	c_{11} (GPa)	c_{13} (GPa)	c_{33} (GPa)	$c_{55}=G_L$ (GPa)	ρ (g cm^{-3})
Mean value (reference)	16.6	9.5	23.1	8.7	15.1	4.7	1.722
Minimum	13.4	6.5	17.6	5.1	9.1	3.3	1.66
Maximum	20.6	12.8	29.6	15.9	25.9	5.5	1.753

mean value of the property S is identical for all δ. The quantity $S(x_3)$ is given by

$$S(x_3) = \frac{(S_m + S_M)}{2} + \delta \times \left(x_3 - \frac{d}{2} \right). \tag{26}$$

The maximal value of δ is also given by Eq. (25), so that all values of $S(x_3)$ are again always comprised between S_m and S_M. Again, the maximal value δ_M of δ is given by Eq. (25).

Gradient of type 2 leads for all magnitudes of δ to a constant value of the spatial average of the gradient.

For both types of spatial variations, five different values of δ regularly distributed between 0 and δ_M are arbitrarily considered in the thickness.

Table II recalls the maximum, minimum, and mean measured values of E_L, E_T, and G_L as given by Dong and Guo (2004). Table II also shows the maximum, minimum, and mean values of the four components (c_{11}, c_{13}, c_{33}, and c_{55}) of the stiffness-tensor $\mathbb{C}$ affecting wave propagation derived from Eq. (A3) of the Appendix.

In Table III, the minimal and maximal values of each variable corresponding to the realistic range of variation obtained (i) by considering the reference values of Table II and (ii) by verifying that the thermodynamical stability conditions are fulfilled. Values resulting from the stability conditions are marked with an asterisk.

In the simulations, the values of S_M and S_m are those reported in Table III.

C. Results and discussion

First of all, the method allows investigating the influence of the fluids on the ultrasonic response. In the case of the characterization of cortical bone, the two fluids f_1 and f_2 are different, which corresponds to an asymmetrical loading (see Sec. IV B): The fluid f_1 has been considered as muscle ($c_{f_1}=1540$ m s^{-1} and $\rho_{f_1}=1.07$ g cm^{-3}) and fluid f_2 as marrow ($c_{f_2}=1450$ m s^{-1} and $\rho_{f_2}=0.9$ g cm^{-3}). The properties of these two fluids are very close to those of water. The frequency spectrum of the reflection coefficient modulus has been plotted for the *in vivo* configuration and compared with the result obtained for a cortical bone plate immersed in water for the ten profiles of mechanical properties (figure not shown). For homogeneous plates as for linearly graded plates, the two curves are very close; however, the modulus of the reflection coefficient at null-frequency is not null and the minimum values are greater than for water but obtained for the same frequency-thickness products. That is why all the following results have been calculated for a cortical bone plate immersed in water.

The reflection coefficient calculated with the Peano series expansion of the matricant is sensitive to the variation of the properties gradient. As we consider that the osteoporosis entails a trabecularization of cortical bone from the endosteal side, the characterization of the gradient of the properties between the endosteal and periosteal regions may be an element of the diagnosis of the osteoporosis progress and of the therapeutic follow-up.

It is known that the gradient of the properties along the cortical thickness is due to the continuous variation of the porosity growing progressively from the periosteal to the endosteal region. From previous work, we know that the porosity influences all the stiffness coefficients (Baron *et al.*, 2007). The frequency spectrum of the reflection coefficient has been plotted for the ten profiles presented in Fig. 3 applied to all the stiffness coefficients implied (c_{11}, c_{13}, c_{33}, and c_{55}) and to the mass density ρ (see Table III). The reflection coefficients have been calculated at an incident angle corresponding to the grazing-angle (critical angle for the longitudinal waves in the bone plate). For the two types of gradients, differences appear between all the gradients and the homogenized plates (corresponding to the maximum value for type 1 and to the average value for type 2) particularly on the location of the extrema values of the reflection coefficient modulus (see Fig. 4).

The increase in the gradient of properties shifts the minimum and maximum values forward high frequency-thickness products. However, the results (see Fig. 4) put on evidence that the behavior of the reflection coefficient modulus is sensibly the same for frequency-thickness products be-

TABLE III. The minimal and maximal values of each variable corresponding to the realistic range of variation obtained (i) by considering the reference values of Table II and (ii) by verifying that the thermodynamical stability conditions are fulfilled. Values resulting from the stability conditions are marked with an asterisk.

Material property S	c_{11} (GPa)	c_{13} (GPa)	c_{33} (GPa)	$c_{55}=G_L$ (GPa)	ρ (g cm^{-3})
Realistic range $[S_m,S_M]$ (reference)	[17.6, 29.6]	[5.1, 11.1*]	[11.8*, 25.9]	[3.3, 5.5]	[1.66, 1.753]

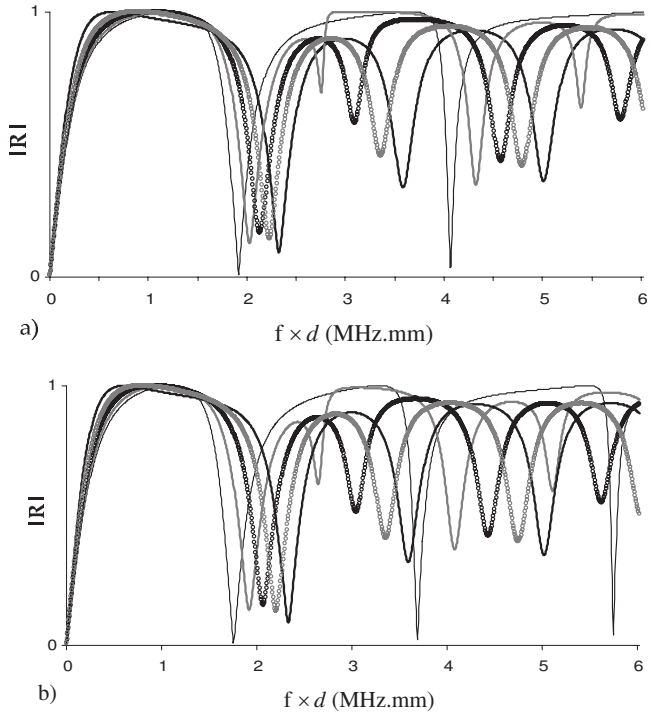


FIG. 4. Frequency spectrum of the reflection coefficient modulus for different properties' variations corresponding to (a) the five profiles of type 1 [Fig. 3(a)] and (b) the five profiles of type 2 [Fig. 3(b)].

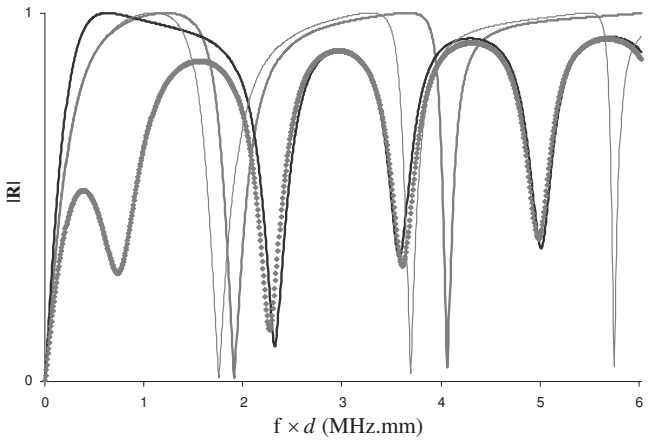


FIG. 5. Frequency spectrum of the reflection coefficient modulus for different properties' variations at grazing-incidence $\theta_1=\theta_c$: for a homogeneous plate whose properties are those at $x_3=0$ at $\theta_c=21.1°$ (thick gray line), for a homogeneous plate whose properties are the average of the properties through the plate at $\theta_c=23.6°$ (thin gray line), for a plate with all its properties linearly varying at $\theta_c=21.1°$ (black line), and for a plate with an affine variation of the c_{11} only at $\theta_c=21.1°$ (gray lozenges).

tween 0 and 1.5 MHz mm. Beyond this value, the behavior is clearly different. At sufficiently high frequency, the wavelength is smaller and more sensitive to the affine variation of the material properties. It is noteworthy that for a heterogeneous waveguide, the minima of the reflection coefficient magnitude do not reach zero (except for null-frequency) but end at a finite value and the changes in phase (not shown) are not so rapid, which means that total transmission does not take place in this situation.

The influence of the variation of each parameter (stiffness coefficients) on the frequency spectrum of the reflection coefficient has been investigated. This analysis has been led for an incidence angle corresponding to the longitudinal wave critical angle in the plate at $x_3=0$. This incidence corresponds to the generation and the propagation of the head wave.

It appears that each of them has an impact on the reflection wave, but the leading term is c_{11}. The frequency spectrum of the reflection coefficient for a varying c_{11} in an affine way is very close to the frequency spectrum of the reflection coefficient obtained for the affine variation of all the material properties (c_{ij} and ρ) and is the most different to the results from homogenized plates (average value or maximum value) compared to the frequency spectrum calculated for the *one-parameter variation* of the other elastic parameters (c_{13}, c_{33}, and c_{55}, and ρ) (see Fig. 5). It is noteworthy that c_{11} is the stiffness coefficient associated with the axial direction and determining the speed of the head wave. So, it seems that the head wave would be the indicator of the c_{11} gradient. It is important to note that for homogenized plates (extremum value or average) the frequency spectrum of the reflection coefficient is really different from that of the plate with continuously varying properties. We infer that the approximation by a homogeneous plate of cortical bone and all the more so for the osteoporotic cortical bone (for which the gradient would be greater) may induce bad interpretation of the ultrasonic response.

The authors do not know any results in literature about the measurement of the variation of the porosity within the cortical thickness. The assumption of an affine gradient is a first step; other gradients may be investigated and the method presented in this article would be applied in the same way for non-affine gradients (Shuvalov *et al.*, 2005).

VI. CONCLUSION AND PERSPECTIVES

Stroh's (1962) sextic plate formalism has been employed for analyzing the leaky Lamb waves in anisotropic heterogeneous plates immersed in fluids. This formalism and especially the polynomial form of the solution [see Eq. (15)] presents several analytical and numerical advantages. First, the low-frequency asymptotics are naturally assessed evaluating only two or three terms in the series (Shuvalov *et al.*, 2005). The information thus collected is of major interest in the analysis of the elastic behavior of waveguides (Baron *et al.*, 2008). Second, the polynomial form makes the numerical evaluation of the solution faster. Indeed, the polynomial coefficients are independent of the frequency, so they are calculated for a fixed slowness value and stored. When the frequency varies, there is no need to recalculate the polynomial coefficients, it comes to a polynomial evaluation whose coefficients are perfectly known, which is time-saving.

The Peano series of the matricant is a method that keeps the continuity of the profiles and so, the authenticity of the problem. One of the key points for methods based on multilayered media to deal with FGM is to relevantly discretize the properties' profiles. The choice of the discretization may lead to some errors especially in the evaluation of the resonances.

This elegant mathematical tool is also very adaptative to different physical problems. In the case studied—a fluid-loaded plane waveguide—the anisotropy, the heterogeneity (continuous or discontinuous variation of properties) and the asymmetric fluid-loading are taken into account without changing the resolution scheme.

Further work needs to be done to relate the results presented in this paper to dispersion curves and propagation of transient and heterogeneous waves in a fluid-loaded continuously heterogeneous waveguide.

Furthermore, from this study, the transient response of a fluid-loaded plate is considered. The frequency spectrum of the reflection coefficient is calculated for incidences between the normal and critical incidences for compression waves in the fluid f_1. Thus, the plate transfer function is calculated in the Fourier domain (x_1-wavenumber, frequency): $\hat{R}(k_1,x_3;\omega)$. A double inverse Fourier transform on (k_1,ω) is applied on $\hat{R}(k_1,x_3;\omega)$ to transform into the space-time domain; the temporal signals can be obtained at different points along the propagation $\mathbf{x}_3$-axis: $R(x_1,x_3;t)$.

Lastly, the formalism presented here is well-adapted to deal with wave propagation in anisotropic tubes with radial property gradients (Shuvalov, 2003). The wave equation keeps the same form as Eq. (12), the state-vector is expressed from the displacement and traction components in the cylindrical basis, and the matrix $\mathbf{Q}$ depends on the radial position r ($\mathbf{Q}=\mathbf{Q}(r)$). In cylindrical homogeneous structures, taking into account an anisotropy more important than transverse isotropy is fussy because there is no analytical solution to the "classical" wave equation (second-order differential equation). Stroh's (1962) formalism, upon which the Peano expansion of the matricant is based, is a promising alternative solution that allows considering altogether the geometry (cylinder), the anisotropy, and the heterogeneity (radial property gradients) of a structure.

APPENDIX: THERMODYNAMICS STABILITY CONDITIONS AND STIFFNESS COEFFICIENTS

Hooke's law is written under the form $\sigma_{ij}=C_{ijk\ell}\epsilon_{k\ell}$ for $(i,j,k,\ell=1,\ldots,3)$, where σ is the stress-tensor, ϵ is the strain-tensor, and $\mathbb{C}$ is the fourth-order stiffness-tensor. In the transversely isotropic case, with $(\mathbf{x}_2,\mathbf{x}_3)$ as isotropic plane, the stiffness-tensor is expressed as a stiffness matrix (using Voigt's notation)

$$\mathbb{C}=\begin{pmatrix} c_{11} & c_{13} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{33} & c_{23} & 0 & 0 & 0 \\ c_{13} & c_{23} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{55} \end{pmatrix}. \tag{A1}$$

We introduce the matrix $\mathbb{S}$, the inverse of the matrix $\mathbb{C}$. It is expressed by

$$\mathbb{S}=\begin{pmatrix} 1/E_L & -\nu_L/E_L & -\nu_L/E_L & 0 & 0 & 0 \\ -\nu_L/E_L & 1/E_T & -\nu_T/E_T & 0 & 0 & 0 \\ -\nu_L/E_L & -\nu_T/E_T & 1/E_L & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/G_T & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/G_L & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/G_L \end{pmatrix}. \tag{A2}$$

with $E_{L,T}$, the longitudinal (L) and transverse (T) Young's moduli; $\nu_{L,T}$, the longitudinal (L) and transverse (T) Poisson's ratios; and $G_{L,T}$, the longitudinal (L) and transverse (T) shear moduli. By inverting Eq. (A1)and identifying it with Eq. (A2), we obtain the following relations:

$$E_L=\frac{c_{11}c_{33}-2c_{13}^2+c_{11}c_{23}}{c_{33}+c_{23}}, \quad \nu_L=\frac{c_{13}}{c_{33}+c_{23}},$$

$$E_T=\frac{c_{11}(c_{33}^2-c_{23}^2)+2c_{13}^2(c_{23}-c_{33})}{c_{11}c_{33}-c_{13}^2}, \quad \nu_T=\frac{c_{11}c_{23}-c_{13}^2}{c_{11}c_{33}-c_{13}^2},$$

$$G_T=c_{44}, \quad G_L=c_{55}. \tag{A3}$$

Knowing the stiffness coefficient values, we can verify if the thermodynamical stability conditions are satisfied as follows:

$$E_L>0, \quad E_T>0, \quad -1<\nu_T<1,$$

$$\frac{(1-\nu_T)}{2}\frac{E_L}{E_T}-\nu_L^2>0, \quad G_L>0. \tag{A4}$$

Baron, C. (**2005**). "Peano expansion of the matricant to study elastic wave propagation in continuously heterogeneous media," University Bordeaux 1 (in French).

Baron, C., Poncelet, O., Shuvalov, A., and Deschamps, M. (**2008**). "Propagation in continuous stratification environment," in *Materials and Acoustics Handbook*, edited by M. Bruneau and C. Potel (ISTE Ltd. and Wiley, London, UK), Chap. 6.

Baron, C., Talmant, M., and Laugier, P. (**2007**). "Effect of porosity on effective diagonal stiffness coefficients (c_{ii}) and anisotropy of cortical at 1 MHz: A finite-difference time domain study," J. Acoust. Soc. Am. **122**, 1810–1817.

Bossy, E., Talmant, M., and Laugier, P. (**2004a**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Bossy, E., Talmant, M., Peyrin, F., Akrout, L., Cloetens, P., and Laugier, P. (**2004b**). "An *in vitro* study of the ultrasonic axial transmission technique at the radius: 1-MHz velocity measurements are sensitive to both mineralization and intracortical porosity," J. Bone Miner. Res. **19**, 1548–1556.

Bousson, V., Meunier, A., Bergot, C., Vicaut, E., Rocha, M. A., Morais, M. H., Laval-Jeantet, A. M., and Laredo, J. D. (**2001**). "Distribution of intracortical porosity in human midfemoral cortex by age and gender," J. Bone Miner. Res. **16**, 1308–1317.

Brekhovskikh, L. M. (**1980**). *Waves in Layered Media* (Academic, New York).

Burlew, M., Madsen, E., Zagzebski, J., Banjavic, R., and Sum, S. (**1980**). "A new ultrasound tissue-equivalent material," Radiology **134**, 517–520.

Camus, E., Talmant, M., Berger, G., and Laugier, P. (**2000**). "Analysis of the axial transmission technique for the assessment of skeletal status," J. Acoust. Soc. Am. **108**, 3058–3065.

Chimenti, D. E., and Nayfeh, A. H. (**1986**). "Anomalous ultrasonic dispersion in fluid-coupled, fibrous composite plates," Appl. Phys. Lett. **49**, 492–493.

Chimenti, D. E., and Nayfeh, A. H. (**1990**). "Ultrasonic reflection and guided waves in fluid-coupled composite laminates," J. Nondestruct. Eval. **9**, 51–69.
Chimenti, D. E., and Rokhlin, S. (**1990**). "Relationship between leaky Lamb modes and reflection coefficient zeroes for a fluid-coupled elastic layer," J. Acoust. Soc. Am. **88**, 1603–1611.
Deschamps, M., and Poncelet, O. (**2000**). "Transient Lamb waves: Comparison between theory and experiment," J. Acoust. Soc. Am. **107**, 3120–3129.
Dickey, J., Maidanik, G., and Uberall, H. (**1995**). "The splitting of dispersion curves for the fluid-loaded plate," J. Acoust. Soc. Am. **98**, 2365–2367.
Dong, X. N., and Guo, X. E. (**2004**). "The dependence of transversely isotropic elasticity of human femoral cortical bone on porosity," J. Biomech. **37**, 1281–1287.
Fiorito, R., Madigosky, W., and Überall, H. (**1979**). "Resonance theory of acoustic waves interacting with an elastic plate," J. Acoust. Soc. Am. **66**, 1857–1866.
Folds, D. L., and Loggins, C. D. (**1977**). "Transmission and reflection of ultrasonic waves in layered media," J. Acoust. Soc. Am. **62**, 1102–1109.
Franklin, H., Danila, E., and Conoir, J.-M. (**2001**). "S-matrix theory applied to acoustic scattering by asymmetrically fluid-loaded elastic isotropic plates," J. Acoust. Soc. Am. **110**, 243–253.
Gantmacher, F. (**1959**). *The Theory of Matrices* (Wiley, New York).
Haskell, N. A. (**1953**). "The dispersion of surface waves on multilayered media," Bull. Seismol. Soc. Am. **43**, 377–393.
Hill, C., Bamber, J., and Ter Haar, G. (**1986**). *Physical Principles of Medical Ultrasonics* (Wiley, New York).
Hosten, B., and Castaings, M. (**2003**). "Surface impedance matrices to model the propagation in multilayered media," Ultrasonics **41**, 501–507.
Kenneth, E. G. (**1982**). "A propagator matrix method for periodically stratified media," J. Acoust. Soc. Am. **73**, 137–142.
Lamb, H. (**1917**). "On waves in an elastic plate," Proc. R. Soc. London Ser. A **93**, 114–128.
Lévesque, D., and Piché, L. (**1992**). "A robust transfer matrix simulation for ultrasonic response of multilayered absorbing media," J. Acoust. Soc. Am. **92**, 452–467.
Macocco, K., Grimal, Q., Naili, S., and Soize, C. (**2006**). "Elastoacoustic model with uncertain mechanical properties for ultrasonic wave velocity prediction; application to cortical bone evaluation," J. Acoust. Soc. Am. **119**, 729–740.
Marin, F., Gonzales-Macias, J., Diez-Perez, A., Palma, S., and Delgado-Rodriguez, M. (**2006**). "Relationship between bone quantitative ultrasound and fractures: A meta-analysis," J. Bone Miner. Res. **21**, 1126–1135.
Nayfeh, A. H., and Chimenti, D. E. (**1988**). "Ultrasonic wave reflection from liquid-coupled orthotropic plates with application to fibrous composites," ASME J. Appl. Mech. **55**, 863–870.
Nayfeh, A. H., and Chimenti, D. E. (**1989**). "Free wave propagation in plates of general anisotropic media," ASME J. Appl. Mech. **56**, 881–886.
Peano, G. (**1888**). "Integration by series of linear differential equations (in French)," Math. Ann. **32**, 450–456.
Pease, M. (**1965**). *Methods of Matrix Algebra* (Academic, New York).
Pitts, L. E., Plona, T. J., and Mayer, W. G. (**1976**). "Theoretical similarities of Rayleigh and Lamb modes of vibration," J. Acoust. Soc. Am. **60**, 374–377.
Rayleigh, J. W. S. (**1885**). "On wave propagating along the plane surface of an elastic solid," Proc. London Math. Soc. **s1-17**, 4–11.
Reilly, D. T., and Burnstein, A. H. (**1974**). "The mechanical properties of cortical bone," J. Bone Jt. Surg., Am. Vol. **56**, 1001–1022.
Rho, J. Y. (**1996**). "An ultrasonic method for measuring the elastic properties of human tibial cortical and cancellous bone," Ultrasonics **34**, 777–783.
Rokhlin, S. I., and Wang, L. (**2002**). "Stable recursive algorithm for elastic wave propagation in layered anisotropic media: Stiffness matrix method," J. Acoust. Soc. Am. **112**, 822–834.
Rose, J. L. (**2002**). "A baseline and vision of ultrasonic guided wave inspection potential," J. Pressure Vessel Technol. **124**, 273–282.
Shuvalov, A. (**2003**). "A sextic formalism for three-dimensional elastodynamics of cylindrically anisotropic radially inhomogeneous materials," Proc. R. Soc. London, Ser. A **459**, 1611–1639.
Shuvalov, A., Poncelet, O., Deschamps, M., and Baron, C. (**2005**). "Long-wavelength dispersion of acoustic waves in transversely inhomogeneous anisotropic plates," Wave Motion **42**, 367–382.
Stroh, A. N. (**1962**). "Steady state problems in anisotropic elasticity," J. Math. Phys. **41**, 77–103.
Thomson, W. T. (**1950**). "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93.
Vlasie, V., and Rousseau, M. (**2004**). "Guided modes in a plane elastic layer with gradually continuous acoustic properties," NDT & E Int. **37**, 633–644.
Wang, L., and Rokhlin, S. I. (**2001**). "Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media," Ultrasonics **39**, 413–424.
Worlton, H. (**1961**). "Experimental confirmation of Lamb waves at megacycle frequencies," J. Appl. Phys. **32**, 967–971.
Youssef, I., and El-Arabawi, H. (**2007**). "Picard iteration algorithm combined with Gauss-Seidel technique for initial value problems," Appl. Math. Comput. **190**, 345–355.

Convolutional perfectly matched layer for elastic second-order wave equation

YiFeng Li and Olivier Bou Matar[a)]
Joint European Laboratory LEMAC, Institut d'Electronique, de Microelectronique et de Nanotechnologie (IEMN-DOAE-UMR CNRS 8520), Ecole Centrale de Lille, 59651 Villeneuve d'Ascq, France

(Received 14 April 2009; revised 25 September 2009; accepted 17 December 2009)

In this work, a method is presented to extend the convolutional perfectly matched layer (C-PML) to simulate acoustic wave propagation in elastic media with a second-order equation formulation. This non-physical layer is used at the computational edge of a finite element method algorithm in frequency domain, and a pseudo-spectral algorithm in time domain, as an absorbing boundary condition (ABC) to truncate unbounded media. Numerical results show that the C-PML ABC attenuates the outgoing surface waves more effectively than classical PML ABC as proposed by Berenger [J. Comput. Phys. **114**, 195–200 (1994)] for electromagnetic waves in the case of oblique incidence, where the PML method suffers from large spurious reflections. Moreover, a modification of the proposed C-PML formulation is also discussed in order to stabilize the absorbing layer in anisotropic solids where numerical instabilities can appear.
 [DOI: 10.1121/1.3290999]

PACS number(s): 43.40.Fz, 43.20.Bi, 43.20.Fn, 43.20.Gp [DF] Pages: 1318–1327

I. INTRODUCTION

Numerical solutions of partial differential equations for wave propagation require the truncation of an unbounded media to fit into computers with a limited memory and computation time. For such problems, absorbing boundary conditions (ABCs) are needed at the truncated boundary to eliminate the reflections from this boundary to the computational domain.

In 1994, a new ABC, called perfectly matched layer (PML), has been introduced by Berenger[1] for electromagnetic waves. It has been proven to be the most robust and efficient technique for the termination of unbounded domain.[2] Since then, the technique has become classical in numerical simulation of wave propagation and has been extended, among others, for acoustic[3] and elastic waves in isotropic[4,5] and anisotropic media.[6] It has been proven that theoretically at the interface between a physical medium and a perfectly matched medium no reflection occurs, and the incident waves from the physical medium are completely absorbed, regardless of their incidence angle and frequency.[1,4] Nevertheless, this property will be lost when a discretization is needed for numerical implementation, especially in the case of oblique incidence. One then needs to optimize the PML parameters in order to decrease parasitic reflections.[7]

Convolutional perfectly matched layer (C-PML), first presented in electromagnetism by Roden and Gedney[8] and applied in the simulation of elastic wave propagation,[9–12] has been shown to improve the behavior of the discrete PML for grazing angle encountered in the case of surface waves. The main advantage of C-PML over the classical PML is that it is based on the unsplit components of the wave field and leads to a more stable scheme. Moreover, it is highly effective at absorbing signals of long time-signature,[13] surface waves,[12] or in elongated domains of calculation.[10]

Classically, C-PML has been introduced in first-order formulation of both electromagnetism and elastodynamics. In this paper, we propose to extend the C-PML absorbing layer to the second-order system describing elastic waves in displacement formulation in anisotropic solids, as it was done for classical split PML.[14] This second-order formulation will be described in frequency and time domains. In frequency domain, this technique is easy to implement in commercial software based on finite element method (FEM).

The efficiency of this second-order perfectly matched layer is demonstrated based on two dimensional (2D) benchmarks both for isotropic and anisotropic solids and for bulk and surface wave propagation. The simulations are realized with the commercially available software COMSOL MULTIPHYSICS in frequency domain and with a pseudo-spectral (PS) method in time domain. In some anisotropic media, numerical instabilities appear in the C-PML, and a modification of the proposed formulation is also discussed in order to stabilize the absorbing layer.

II. SECOND-ORDER FORMULATION OF ELASTIC WAVE PROPAGATION

Consider the propagation of 2D plane strain elastic waves in an anisotropic elastic solid medium. With Einstein's convention of summation, the equation of motion can be written as

$$\frac{\partial^2 u_i}{\partial t^2} = \frac{1}{\rho_0}\frac{\partial \tau_{ij}}{\partial x_j}, \tag{1}$$

where $i,j=1,2$, ρ_0 is the density, x_j are the components of the position vector, t is the time, u_i are the components of the particle displacement vector, and τ_{ij} are the components of

a)Author to whom correspondence should be addressed. Electronic mail: olivier.boumatar@iemn.univ-lille1.fr

the stress tensor. For a linear elastic solid, the constitutive relation is given by Hooke's law:

$$\tau_{ij} = C_{ijkl}\varepsilon_{kl}, \tag{2}$$

where C_{ijkl} are the elastic constants, and the linear approximation of the strain tensor $\boldsymbol{\varepsilon}$ is

$$\varepsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right). \tag{3}$$

In the case of a transverse isotopic medium, the following second-order system of equation is obtained:

$$\frac{\partial^2 u_1}{\partial t^2} = \frac{1}{\rho_0}\left(\frac{\partial \tau_{11}}{\partial x} + \frac{\partial \tau_{12}}{\partial y}\right), \tag{4a}$$

$$\frac{\partial^2 u_2}{\partial t^2} = \frac{1}{\rho_0}\left(\frac{\partial \tau_{12}}{\partial x} + \frac{\partial \tau_{22}}{\partial y}\right), \tag{4b}$$

$$\tau_{11} = C_{11}\frac{\partial u_1}{\partial x} + C_{12}\frac{\partial u_2}{\partial y}, \tag{4c}$$

$$\tau_{22} = C_{12}\frac{\partial u_1}{\partial x} + C_{22}\frac{\partial u_2}{\partial y}, \tag{4d}$$

$$\tau_{12} = C_{66}\left(\frac{\partial u_1}{\partial y} + \frac{\partial u_2}{\partial x}\right). \tag{4e}$$

Here we have considered that $x_1 = x$ and $x_2 = y$. This system will be used as the starting point in the remainder of the paper.

III. C-PML FOR ELASTIC WAVE EQUATIONS IN FREQUENCY DOMAIN

Here, the methodology used for the introduction of C-PML zones for axisymmetric active solid media[9] is used for the system of Eq. (4). First, taking Fourier transform of the system, it is rewritten in the frequency domain. Then, the following complex coordinate transformation[15] is used:

$$\tilde{x} = \int_0^x s_x(x')dx', \tag{5a}$$

$$\tilde{y} = \int_0^y s_y(y')dy', \tag{5b}$$

where s_x and s_y are the complex frequency shifted (CFS) stretched-coordinate metrics proposed by Kuzuoglu and Mittra:[16]

$$s_x(x) = \kappa_x(x) + \frac{\sigma_x(x)}{\alpha_x(x) + j\omega}, \tag{6a}$$

$$s_y(y) = \kappa_y(y) + \frac{\sigma_y(y)}{\alpha_y(y) + j\omega}. \tag{6b}$$

α_x, σ_x, α_y, and σ_y are assumed to be positive and real, and κ_x and κ_y are real and ≥ 1. The $\sigma_{x,y}$ and $\kappa_{x,y}$ are the so-called[2] attenuation factor used for the attenuation of propagating waves and scaling factor used for the attenuation of evanescent waves, respectively. The $\alpha_{x,y}$ are frequency-dependent terms that implement a Butterworth-type filter in the layer. The originally split PML, introduced for isotropic elastic wave propagation by Chew and Liu[4] and for anisotropic solid by Collino and Tsogka,[6] is retrieved imposing $\alpha_{x,y} = 0$ and $\kappa_{x,y} = 1$. Using the complex coordinate variables $\tilde{x}$ and $\tilde{y}$ to replace x and y in Eqs. (4a)–(4e) and noting that $\partial/\partial\tilde{x} = (1/s_x)\,\partial/\partial x$ and $\partial/\partial\tilde{y} = (1/s_y)\,\partial/\partial y$, we obtain the following frequency domain equations in Cartesian coordinates:

$$-\omega^2\rho_0\hat{u}_1 = \frac{1}{s_x}\frac{\partial \hat{\tau}_{11}}{\partial x} + \frac{1}{s_y}\frac{\partial \hat{\tau}_{12}}{\partial y}, \tag{7a}$$

$$-\omega^2\rho_0\hat{u}_2 = \frac{1}{s_x}\frac{\partial \hat{\tau}_{12}}{\partial x} + \frac{1}{s_y}\frac{\partial \hat{\tau}_{22}}{\partial y}, \tag{7b}$$

$$\hat{\tau}_{11} = C_{11}\frac{1}{s_x}\frac{\partial \hat{u}_1}{\partial x} + C_{12}\frac{1}{s_y}\frac{\partial \hat{u}_2}{\partial y}, \tag{7c}$$

$$\hat{\tau}_{22} = C_{12}\frac{1}{s_x}\frac{\partial \hat{u}_1}{\partial x} + C_{22}\frac{1}{s_y}\frac{\partial \hat{u}_2}{\partial y}, \tag{7d}$$

$$\hat{\tau}_{12} = C_{66}\left(\frac{1}{s_y}\frac{\partial \hat{u}_1}{\partial y} + \frac{1}{s_x}\frac{\partial \hat{u}_2}{\partial x}\right), \tag{7e}$$

where $\hat{u}$ represents the Fourier transform of the variable u.

In order to implement this C-PML in a commercial FEM software (COMSOL MULTIPHYSICS), the resulting second-order C-PML wave equations are interpreted as an anisotropic medium, as it has already been done for PML.[17] Multiplying Eq. (7) by $s_x s_y$ and introducing new stress tensor $\hat{\tau}'_{ij}$ and density ρ'_0 (given by $\rho_0 s_x s_y$), we get the following equations:

$$-\omega^2\rho'_0\hat{u}_1 = \frac{\partial \hat{\tau}'_{11}}{\partial x} + \frac{\partial \hat{\tau}'_{12}}{\partial y}, \tag{8a}$$

$$-\omega^2\rho'_0\hat{u}_2 = \frac{\partial \hat{\tau}'_{21}}{\partial x} + \frac{\partial \hat{\tau}'_{22}}{\partial y}, \tag{8b}$$

$$\hat{\tau}'_{11} = \hat{\tau}_{11}s_y = C_{11}\frac{s_y}{s_x}\frac{\partial \hat{u}_1}{\partial x} + C_{12}\frac{\partial \hat{u}_2}{\partial y}, \tag{8c}$$

$$\hat{\tau}'_{22} = \hat{\tau}_{22}s_x = C_{12}\frac{\partial \hat{u}_1}{\partial x} + C_{22}\frac{s_x}{s_y}\frac{\partial \hat{u}_2}{\partial y}, \tag{8d}$$

$$\hat{\tau}'_{12} = \hat{\tau}_{12}s_x = C_{66}\left(\frac{s_x}{s_y}\frac{\partial \hat{u}_1}{\partial y} + \frac{\partial \hat{u}_2}{\partial x}\right), \tag{8e}$$

$$\hat{\tau}'_{21} = \hat{\tau}_{12}s_y = C_{66}\left(\frac{\partial \hat{u}_1}{\partial y} + \frac{s_y}{s_x}\frac{\partial \hat{u}_2}{\partial x}\right). \tag{8f}$$

This system of equation corresponds to the propagation of elastic waves in a "fictious" anisotropic medium and can be written in the matrix form as

$$\begin{bmatrix} \hat{\tau}'_{11} \\ \hat{\tau}'_{22} \\ \hat{\tau}'_{12} \\ \hat{\tau}'_{21} \end{bmatrix} = \begin{bmatrix} C'_{11} & C_{12} & 0 & 0 \\ C_{12} & C'_{22} & 0 & 0 \\ 0 & 0 & C'_{66} & C_{66} \\ 0 & 0 & C_{66} & C''_{66} \end{bmatrix} \begin{bmatrix} u_{1,x} \\ u_{2,y} \\ u_{1,y} \\ u_{2,x} \end{bmatrix}, \tag{9}$$

with $u_{1,x}=\partial\hat{u}_1/\partial x$, $u_{2,y}=\partial\hat{u}_2/\partial y$, $u_{1,y}=\partial\hat{u}_1/\partial y$, and $u_{2,x}=\partial\hat{u}_2/\partial x$. The new effective elastic stiffnesses are $C'_{11}=C_{11}s_y/s_x$, $C'_{22}=C_{22}s_x/s_y$, $C'_{66}=C_{66}s_x/s_y$, and $C''_{66}=C_{66}s_y/s_x$. It should be noted that this fictious anisotropic medium has a nonsymmetric stress tensor ($\hat{\tau}_{ij}\neq\hat{\tau}_{ji}$ when $i\neq j$), and the complex-valued tensor of elastic constants conserves minor symmetry properties, but not the major one.

We can easily extend this description of C-PML in anisotropic solid to three dimensions. In this case, the general form of the propagation of elastic waves can be described as

$$-\omega^2\rho'_0\hat{u}_i=\frac{\partial\hat{\tau}'_{ij}}{\partial x_j}, \tag{10a}$$

$$\hat{\tau}'_{ij}=C'_{ijkl}\frac{\partial\hat{u}_l}{\partial x_k}, \tag{10b}$$

where $i,j,k,l=1$, 2, or 3. The effective elastic tensor C' and the density ρ'_0 are given by

$$C'_{ijkl}=C_{ijkl}\frac{s_xs_ys_z}{s_is_k}, \tag{11a}$$

$$\rho'_0=\rho_0s_xs_ys_z. \tag{11b}$$

The effective constants already obtained for 2D situation can be easily derived from Eq. (11a) by considering that $s_z=1$ and $i,j,k,l=1$ or 2. Moreover, this fictious anisotropic interpretation of C-PML or PML can be extended to piezoelectric solids.[18–20]

IV. C-PML FOR ELASTIC WAVE EQUATIONS IN TIME DOMAIN

Now, to be able to obtain a C-PML formulation in time domain, the resulting equations are transformed back to time domain by inverse Fourier transform. Due to the frequency dependence of the CFS stretched-coordinate metrics, a convolution appears in the resulting equations, as shown, for example, for Eq. (7a):

$$\frac{\partial^2u_1}{\partial t^2}=\frac{1}{\rho_0}\left[F^{-1}\left[\frac{1}{s_x}\right]\otimes\frac{\partial\tau_{11}}{\partial x}+F^{-1}\left[\frac{1}{s_y}\right]\otimes\frac{\partial\tau_{12}}{\partial y}\right], \tag{12}$$

where $\otimes$ and $F^{-1}[\cdot]$ are, respectively, the convolution and inverse Fourier transform operators. In order to eliminate the convolutions appearing in Eq. (12), we use the methodology introduced by Roden and Gedney[8] in electromagnetism and extended by Bou Matar *et al.*[9] for elastic wave propagation in active (or nonlinear) media by introducing memory variables. The time evolution of each of these memory variables is realized by a first-order differential equation. The obtained result can be rewritten as

$$\frac{\partial^2u_1}{\partial t^2}=\frac{1}{\rho_0}\left(\frac{1}{\kappa_x}\frac{\partial\tau_{11}}{\partial x}+\frac{1}{\kappa_y}\frac{\partial\tau_{12}}{\partial y}+\frac{A_x}{\kappa_x}+\frac{B_y}{\kappa_y}\right), \tag{13}$$

where the memory variables A_x and B_y are given by

$$\frac{\partial A_x}{\partial t}=-\delta_x\frac{\partial\tau_{11}}{\partial x}-\beta_xA_x, \tag{14a}$$

$$\frac{\partial B_y}{\partial t}=-\delta_y\frac{\partial\tau_{12}}{\partial y}-\beta_yB_y, \tag{14b}$$

with $\delta_{x,y}=\sigma_{x,y}/\kappa_{x,y}$ and $\beta_{x,y}=\sigma_{x,y}/\kappa_{x,y}+\alpha_{x,y}$. Making the same calculation for Eqs. (7b)–(7e), we obtain the following system equations of C-PML in time domain for elastic wave propagation in anisotropic solid:

$$\frac{\partial^2u_1}{\partial t^2}=\frac{1}{\rho_0}\left(\frac{1}{\kappa_x}\frac{\partial\tau_{11}}{\partial x}+\frac{1}{\kappa_y}\frac{\partial\tau_{12}}{\partial y}+\frac{A_x}{\kappa_x}+\frac{B_y}{\kappa_y}\right), \tag{15a}$$

$$\frac{\partial^2u_2}{\partial t^2}=\frac{1}{\rho_0}\left(\frac{1}{\kappa_x}\frac{\partial\tau_{12}}{\partial x}+\frac{1}{\kappa_y}\frac{\partial\tau_{22}}{\partial y}+\frac{C_x}{\kappa_x}+\frac{D_y}{\kappa_y}\right), \tag{15b}$$

$$\tau_{11}=\frac{C_{11}}{\kappa_x}\frac{\partial u_1}{\partial x}+\frac{C_{12}}{\kappa_y}\frac{\partial u_2}{\partial y}+\frac{C_{11}}{\kappa_x}E_x+\frac{C_{12}}{\kappa_y}F_y, \tag{15c}$$

$$\tau_{22}=\frac{C_{12}}{\kappa_x}\frac{\partial u_1}{\partial x}+\frac{C_{22}}{\kappa_y}\frac{\partial u_2}{\partial y}+\frac{C_{12}}{\kappa_x}E_x+\frac{C_{22}}{\kappa_y}F_y, \tag{15d}$$

$$\tau_{12}=\frac{C_{66}}{\kappa_x}\frac{\partial u_2}{\partial x}+\frac{C_{66}}{\kappa_y}\frac{\partial u_1}{\partial y}+\frac{C_{66}}{\kappa_x}G_x+\frac{C_{66}}{\kappa_y}H_y, \tag{15e}$$

where the memory variables C_x, D_y, E_x, F_y, G_x, and H_y are obtained by the first-order differential equations similar to the one derived for A_x and B_y in Eq. (14). The memory variables will be zero outside the C-PML zones, so the first-order differential equations of memory variables need only to be solved in a small part of the calculation domain. After introducing Eqs. (15c)–(15e) into Eqs. (15a) and (15b), the resulting system of wave equations can be written as

$$\rho_0\frac{\partial^2u_1}{\partial t^2}-\left[\frac{\partial}{\partial x}\left(C_{11}\frac{\partial u_1}{\partial x}+C_{12}\frac{\partial u_2}{\partial y}\right)+\frac{\partial}{\partial y}\left(C_{66}\frac{\partial u_2}{\partial x}+C_{66}\frac{\partial u_1}{\partial y}\right)\right]=f_1, \tag{16a}$$

$$\rho_0\frac{\partial^2u_2}{\partial t^2}-\frac{\partial}{\partial x}\left(C_{66}\frac{\partial u_2}{\partial x}+C_{66}\frac{\partial u_1}{\partial y}\right)+\frac{\partial}{\partial y}\left(C_{12}\frac{\partial u_1}{\partial x}+C_{22}\frac{\partial u_2}{\partial y}\right)=f_2, \tag{16b}$$

where f_1 and f_2 are

$$f_1 = \left[-\frac{\kappa'_x}{\kappa_x}\frac{\partial}{\partial x}\left(\frac{C_{11}}{\kappa_x}\frac{\partial u_1}{\partial x} + \frac{C_{12}}{\kappa_y}\frac{\partial u_2}{\partial y} + \frac{C_{11}}{\kappa_x}E_x + \frac{C_{12}}{\kappa_y}F_y\right) - \frac{\kappa'_y}{\kappa_y}\frac{\partial}{\partial y}\left(\frac{C_{66}}{\kappa_x}\frac{\partial u_2}{\partial x} + \frac{C_{66}}{\kappa_y}\frac{\partial u_1}{\partial y} + \frac{C_{66}}{\kappa_x}G_x + \frac{C_{66}}{\kappa_y}H_y\right) - \frac{\partial}{\partial x}\left(\frac{\kappa'_x C_{11}}{\kappa_x}\frac{\partial u_1}{\partial x} + \frac{\kappa'_y C_{12}}{\kappa_y}\frac{\partial u_2}{\partial y} - \frac{C_{11}}{\kappa_x}E_x - \frac{C_{12}}{\kappa_y}F_y\right) - \frac{\partial}{\partial y}\left(\frac{\kappa'_x C_{66}}{\kappa_x}\frac{\partial u_2}{\partial x} + \frac{\kappa'_y C_{66}}{\kappa_y}\frac{\partial u_1}{\partial y} - \frac{C_{66}}{\kappa_x}G_x - \frac{C_{66}}{\kappa_y}H_y\right) + \frac{A_x}{\kappa_x} + \frac{B_y}{\kappa_y}\right], \tag{16c}$$

$$f_2 = \left[-\frac{\kappa'_x}{\kappa_x}\frac{\partial}{\partial x}\left(\frac{C_{66}}{\kappa_x}\frac{\partial u_2}{\partial x} + \frac{C_{66}}{\kappa_y}\frac{\partial u_1}{\partial y} + \frac{C_{66}}{\kappa_x}G_x + \frac{C_{66}}{\kappa_y}H_y\right) - \frac{\kappa'_y}{\kappa_y}\frac{\partial}{\partial y}\left(\frac{C_{12}}{\kappa_x}\frac{\partial u_1}{\partial x} + \frac{C_{22}}{\kappa_y}\frac{\partial u_2}{\partial y} + \frac{C_{12}}{\kappa_x}E_x + \frac{C_{22}}{\kappa_y}F_y\right) - \frac{\partial}{\partial x}\left(\frac{\kappa'_x C_{66}}{\kappa_x}\frac{\partial u_2}{\partial x} + \frac{\kappa'_y C_{66}}{\kappa_y}\frac{\partial u_1}{\partial y} - \frac{C_{66}}{\kappa_x}G_x - \frac{C_{66}}{\kappa_y}H_y\right) - \frac{\partial}{\partial y}\left(\frac{\kappa'_x C_{12}}{\kappa_x}\frac{\partial u_1}{\partial x} + \frac{\kappa'_y C_{22}}{\kappa_y}\frac{\partial u_2}{\partial y} - \frac{C_{12}}{\kappa_x}E_x - \frac{C_{22}}{\kappa_y}F_y\right) + \frac{C_x}{\kappa_x} + \frac{D_y}{\kappa_y}\right], \tag{16d}$$

where we have introduced the notation $\kappa'_{x,y}=\kappa_{x,y}-1$. The formulation of C-PML given by Eq. (16) has been introduced in order to facilitate the implementation of the resulting second-order C-PML wave equations in commercial finite element packages and in our homemade PS code.

It is important to note that in time domain the number of equation needed to be solved increases in the C-PML zones. This can considerably increase the burden of calculation. Nevertheless, when active (or nonlinear) media are considered, then time domain formulations are needed,[9] and Eq. (16) has to be used. Moreover, the number of auxiliary variables in the unsplit and convolutional PML is less than the one of the classical split PML. In the considered case, elastic wave propagation in a 2D anisotropic solid, eight auxiliary variables are needed for the unsplit PML and ten for the split PML. In both cases, these auxiliary variables need to be stored in addition to the total field variables: u_1 and u_2 for the proposed CPML or u_1, u_2, τ_{11}, τ_{22}, and τ_{12} for the classical split PML. In the classical PML case, one can choose either to store or not the total field. In this last case, the memory storage is reduced but the computation time is increased because one then needs to compute the total field variables as the sum of two splitted auxiliary variables several times in each iteration of the time loop.

V. NUMERICAL TESTS

In this section, the excellent absorbing behaviors of both formulations (frequency and time domains) are demonstrated. The frequency domain formulation has been implemented in a commercial FEM software (COMSOL MULTIPHYSICS), and a PS code has been developed for the time domain formulation. The choice of these numerical schemes has been motivated by the compromise in obtaining both efficient and accurate methods for the two different formulations.

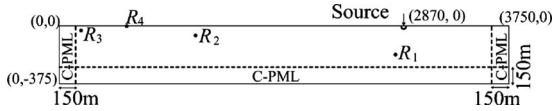


FIG. 1. Schematic of the model used in the example of surface wave in a homogeneous isotropic medium. The C-PML terminations are positioned at the left, right, and bottom sides of the simulation domain. The top boundary is a stress free boundary. The source and four receiver (R_1, R_2, R_3, and R_4) positions are also depicted.

A. Surface wave in a homogeneous isotropic medium

In order to demonstrate the improvement induced by the use of C-PML, instead of classical PML, numerical simulations of elastic wave propagation in a half space of homogeneous isotropic medium, similar to the ones proposed by Drossaert and Giannopoulos,[11] are presented. The sketch of the used elongated calculation domain (3750 m width and 375 m height, including the C-PMLs and corresponding to 200×20 grid elements) is given in Fig. 1, where the origin (0, 0) is at the left upper corner. The properties of the elastic medium have been chosen as follows: the density is $\rho_0=2000$ kg m^{-3}, and the Lame constants are $\lambda=0.6\times 10^9$ and $\mu=0.3\times 10^9$ N/m^2. A directional point source, acting as a force perpendicular to the upper free surface, is located at (2870.0, 0.0). Four receiving positions R_1, R_2, R_3, and R_4 are chosen at the following locations: (2800.0, −187.5), (1120.0, −42.5), (168.75, −18.75), and (562.5, 0), respectively. The choice of these receiving positions has been realized in order to demonstrate typical behavior of the C-PML. The temporal variation of the point source is a Ricker wavelet given by

$$f(t) = (0.5 + a_1(t-t_D)^2)e^{-a_1(t-t_D)^2}, \tag{17}$$

where $t_D=1.0$ s is the source delay time, $a_1=(\pi f_c)^2$, and $f_c=1.0$ Hz is the central frequency.

In the stretched-coordinate metrics, the following spatial coordinate dependences are used for the parameters of the C-PMLs in the x direction:

$$\kappa_x = 1 + \kappa_{\max}\left(\frac{x-x_0}{d_x}\right)^{n_1}, \tag{18a}$$

$$\sigma_x = \sigma_{\max}\left(\frac{x-x_0}{d_x}\right)^{n_1+n_2}, \tag{18b}$$

$$\alpha_x = \alpha_{\max}\left(\frac{d_x - x + x_0}{d_x}\right)^{n_3}, \tag{18c}$$

with $\alpha_{\max}=2\pi f_\alpha$ and

$$\sigma_{\max} = (1+n_1+n_2)\sqrt{(\lambda+2\mu)/\rho_0}\log(1/R_0)/(2d_x), \tag{19}$$

where R_0 is the reflection coefficient from the C-PML we want to obtain, and x_0 and d_x are, respectively, the starting position and thickness of the C-PML. κ_y, σ_y, and α_y have the

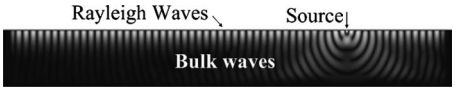


FIG. 2. Displacement amplitude snapshot at frequency 3 Hz for the model depicted in Fig. 1, showing the propagation and absorption in the C-PML terminations of both bulk and Rayleigh waves.

same form as κ_x, σ_x, and α_x, respectively, just replacing x by y. f_α is a characteristic frequency of the C-PML (here $f_\alpha = 10^{-2}$ Hz). It corresponds to a cut off frequency.[21] Indeed, when $f \gg f_\alpha$ the complex frequency shifted coefficients $\alpha_{x,y}$ are negligible, and the attenuation in the C-PML is the same as in the PML. On the other hand, when $f \ll f_\alpha$ the stretch of the coordinates is real and the medium does not absorb propagating waves, but only the evanescent waves. The spatial dependences used in Eq. (18) are the same as the ones classically proposed for PML applied to electromagnetic waves.[2] $\sigma_{\max}$ given by Eq. (19) corresponds to the optimal value proposed in Ref. 6. In the considered cases we use $\kappa_{\max}=0$. Moreover, $\kappa_{x,y}=1$ and $\sigma_{x,y}=0$ in the computational domain, outside the C-PML zone. In the presented results, a C-PML thickness of eight grid elements, both in the left, right, and bottom boundaries, $R_0=5\ \ 10^{-5}$, $n_1=3$, $n_2=0$, and $n_3=1$ has been chosen. We remind also that the classical PML case corresponds to $\alpha_{\max}=0$.

Here, Eq. (8) has been solved with COMSOL MULTIPHYSICS FEM software in frequency domain. Indeed, as demonstrated by Castaings *et al.*,[22] very efficient simulations of linear pulse propagation in solids can be obtained with only a limited number of frequency calculations. The time evolution of the point source [Eq. (17)] is first Fourier transform, and Eq. (8) is then solved for 40 frequencies equally spaced in the source spectrum. Second-order Lagrange elements have been used in all the FEM calculations. The result obtained at one frequency (3 Hz) is shown in Fig. 2, where both bulk waves in the physical domain and Rayleigh waves on the free surface boundary can be seen. In Fig. 3, the time evolutions of the horizontal [(a) and (b)] and vertical [(c) and (d)] displacements reconstructed by inverse Fourier transform of the 40 frequency responses at the fourth receiver, positioned at the free surface, in the case of both C-PML (dotted line) and PML (dashed line) layers for absorbing the surface wave are displayed and compared to an analytical solution (solid line). The FORTRAN code EX2DDIR of Berg *et al.*[23] has been used to compute this exact solution of the 2D response from a vertical directional point source in an isotropic elastic half space with a free surface. The obtained results clearly demonstrated the increased efficiency of the C-PML in comparison with the PML in order to absorb the Rayleigh wave, particularly when a zoom [Figs. 3(b) and 3(d)] is made around the end of this surface wave. This is in perfect accordance with previous results obtained with the C-PML first-order velocity-stress formulation implementation.[10–12]

The time evolutions of the horizontal [(a), (c), and (e)] and vertical [(b), (d), and (f)] displacements for the three other receivers (R_1, R_2, and R_3) in the case of FEM simulation with C-PML (dotted line) are plotted in Fig. 4 and compared, as before, with a FEM simulation with PML (dashed

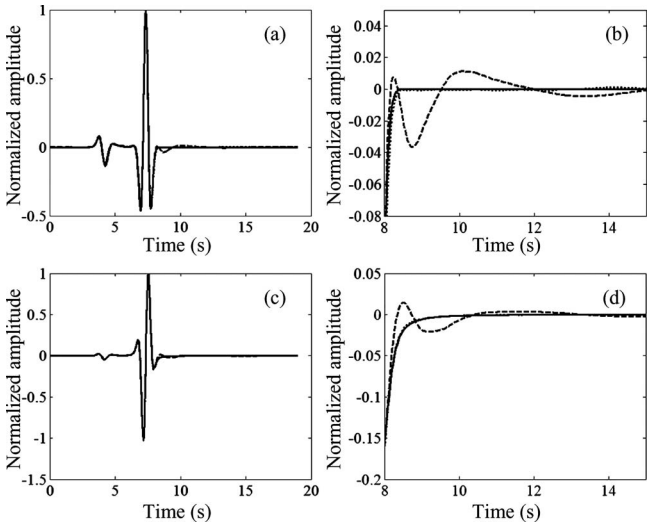


FIG. 3. Time evolution of the horizontal u_1 (a) and the vertical u_2 (c) components of the displacement vector at the fourth receiver of the analytical solution of the problem (solid line) and the numerical solution with PML (dashed line) and C-PML (dotted line). (b) and (d) are zoom of (a) and (c), respectively, showing the benefit of using C-PML instead of PML.

line) and an analytical solution (solid line). As in Fig. 3, it clearly appears in Fig. 4 that the use of C-PML greatly improves the surface wave absorption efficiency of the absorbing layer (see R_2 signal, where spurious oscillations of the horizontal displacement component are observed in the case

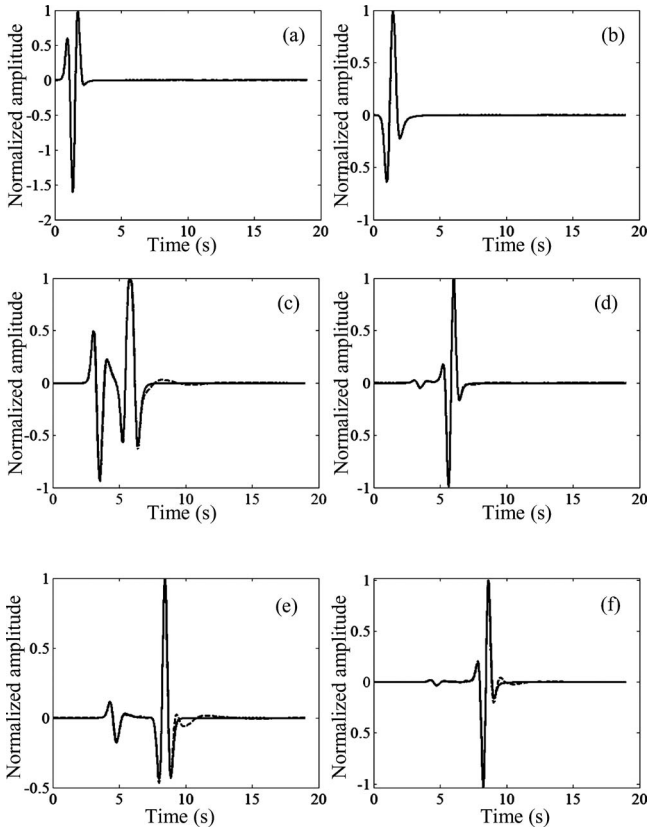


FIG. 4. Time evolution of the horizontal u_1 (left) and the vertical u_2 (right) components of the displacement vector at the first (top), second (middle), and third (bottom) receivers of the analytical solution of the problem (solid line) and the numerical solution with PML (dashed line) and C-PML (dotted line).

of PML), even if the receiver is positioned close to the absorbing layer, as for the R_3 receiver. It is linked to evanescent part of the Rayleigh wave in the bulk direction which interacts over a long distance with the PML at the bottom. In the case of the first receiver, placed in the bulk of the sample where the contribution of surface wave is negligible, the C-PML and PML have the same absorbing efficiency, and the overall agreement with the analytical solution is almost perfect.

B. Bulk wave in a homogeneous anisotropic medium

The PS code described in Ref. 9 has been modified in order to solve the second-order wave equation system of Eq. (16), in place of the classical first-order stress-particle velocity formulation used in finite difference time domain elastic wave propagation simulations. The forward integration in time, realized in our previous PS code by a staggered fourth-order Adams–Bashforth method, is here made with the use of a second-order finite difference method.

Simulations of pulse propagation in anisotropic solids have been made in order to demonstrate the efficiency of the proposed time domain C-PML formulation for second-order wave equations. The properties, density and elasticity coefficients, of the three different orthotropic materials used in the time domain simulations are shown in Table I. The choice of the material name (I, III, and V) has been made, following Bécache *et al.*,[24] in order to simplify the comparison of our results with previous ones.[24–26] In all cases, a 25 cm × 25 cm portion of an infinite solid has been discretized on a 128 × 128 element grid (including the ten element C-PML placed on each side), and a 5 ns time step was used. The following source term $f(x,y,t)$ is added to f_1 in the right hand side of Eq. (16a):

TABLE I. Properties of the orthotropic materials used in the time domain simulations.

Material	ρ (kg/m^3)	C_{11} (GPa)	C_{22} (GPa)	C_{12} (GPa)	C_{66} (GPa)
I	4000	40	200	38	20
III	4000	40	200	75	20
V	4000	300	60	99	15

$$f(x,y,t) = (0.5 + a_1(t - t_D)^2)e^{-a_1(t-t_D)^2}\frac{e^{-7[(x^2+y^2)/r_0^2]}}{r_0^2}, \quad (20)$$

where $a_1 = (\pi f_c)^2$, $f_c = 150$ kHz is the central frequency, $t_D = 1/f_c$ is the source delay time, and $r_0 = 5$ mm. It corresponds to a Gaussian spatial distribution around the (0,0) point which is placed at the center in all the simulations and to a Ricker wavelet time evolution. The same spatial coordinate dependences [Eq. (18)] are used for the C-PML parameters with now $\alpha_{\max} = 20\pi$, $n_1 = 2$, $n_2 = 1$, $n_3 = 1$, and

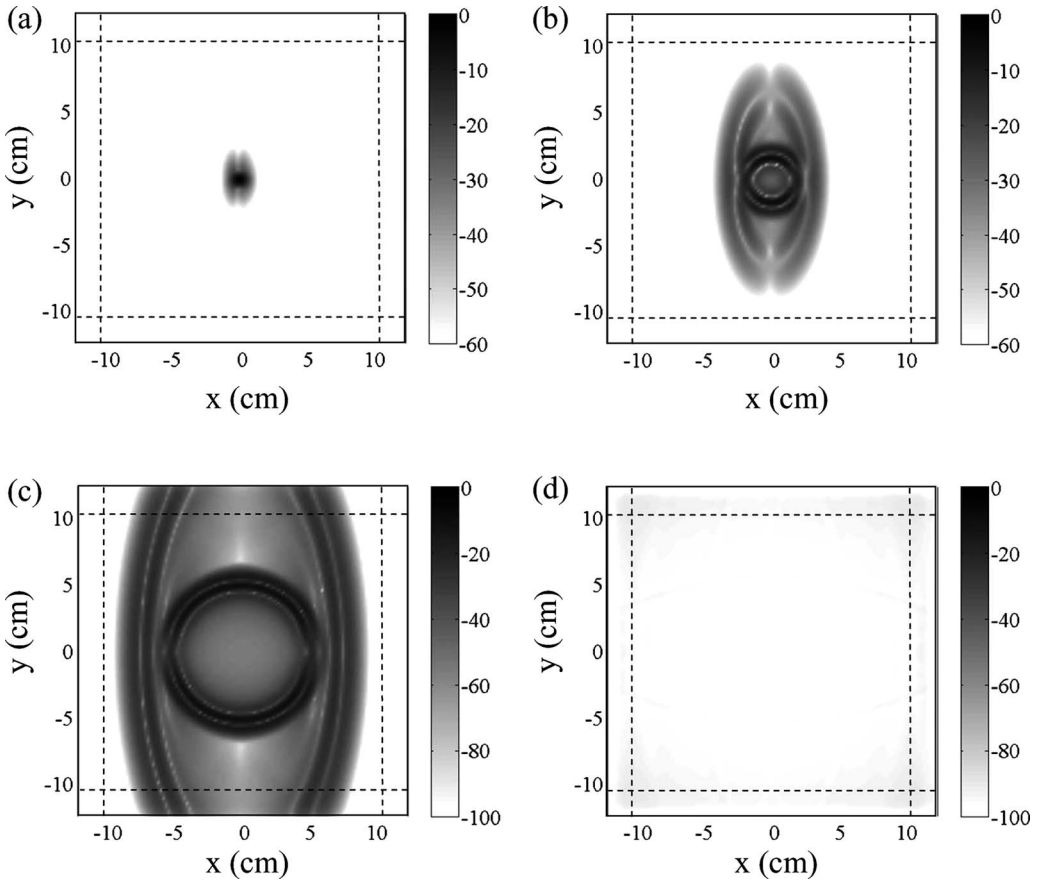


FIG. 5. Snapshots of propagation of the displacement magnitude in an orthotropic elastic medium, model I, at (a) $t = 5$ μs, (b) $t = 15$ μs, (c) $t = 30$ μs, and (d) $t = 125$ μs. The snapshots are in dB scale with a reference displacement amplitude of 10 nm.

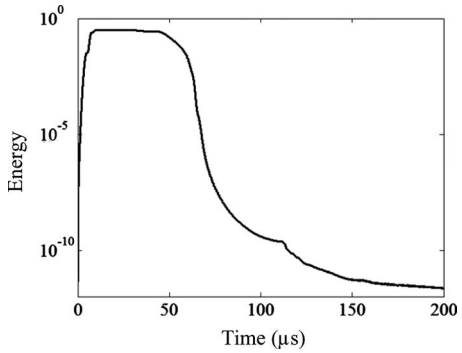


FIG. 6. Energy decay in the physical domain for the orthotropic elastic medium, model I, in the same conditions as the ones used to obtain the snapshots of Fig. 5.

$$\sigma_{\max} = (1 + n_1 + n_2)\sqrt{C_{11}/\rho_0}\log(1/R_0)/(2d_x), \tag{21}$$

where $R_0 = 10^{-12}$.

In Fig. 5, snapshots of the propagation of the amplitude of the displacement in medium I are displayed for four different times (5, 15, 30, and 125 μs). Here and in all the following figures, the snapshots presented are in dB scale with a reference displacement of 10 nm. The overall absorption of the C-PML is excellent and the level of spurious reflection is very small, only a −90 dB reflected pulse can be seen at the later time [Fig. 5(d)]. In order to see more quantitatively the quality of the C-PML, the total energy E at each time is computed according to the following expression:

$$E = \frac{1}{2}\int_\Omega \rho_0 \|\boldsymbol{v}\|^2 d\Omega + \frac{1}{2}\int_\Omega \tau_{ij}\varepsilon_{ij} d\Omega, \tag{22}$$

where Ω is the volume corresponding to the physical domain, and $\|\mathbf{v}\|$ is the magnitude of the velocity vector. Figure 6 displays the energy decay in the physical domain for the proposed C-PML model. The wave field has left the physical domain at about 60 μs here. This figure not only confirms that at later time the energy has decayed by a factor of 10^{12} but also demonstrates that no significant spurious reflection appears before.

Snapshots of the propagation of the amplitude of the displacement in medium III are displayed for four different times (5, 15, 30, and 50 μs) in Fig. 7. As all the other PML or C-PML implementations, some instabilities appear [Figs. 7(c) and 7(d)] in the absorbing layers where, as explained by Bécache *et al.*,[24] one of the incident elastic waves arrives with the components of the group velocity and the slowness vector, in the C-PML direction, of opposite signs. These instabilities are one of the major limitations in the use of PML for elastic waves propagating in anisotropic solids and in plates, even in the case of an isotropic medium.[27] In the case of elastic plate absorbing layers, solutions have been proposed in order to overcome this limitation, but at the expense

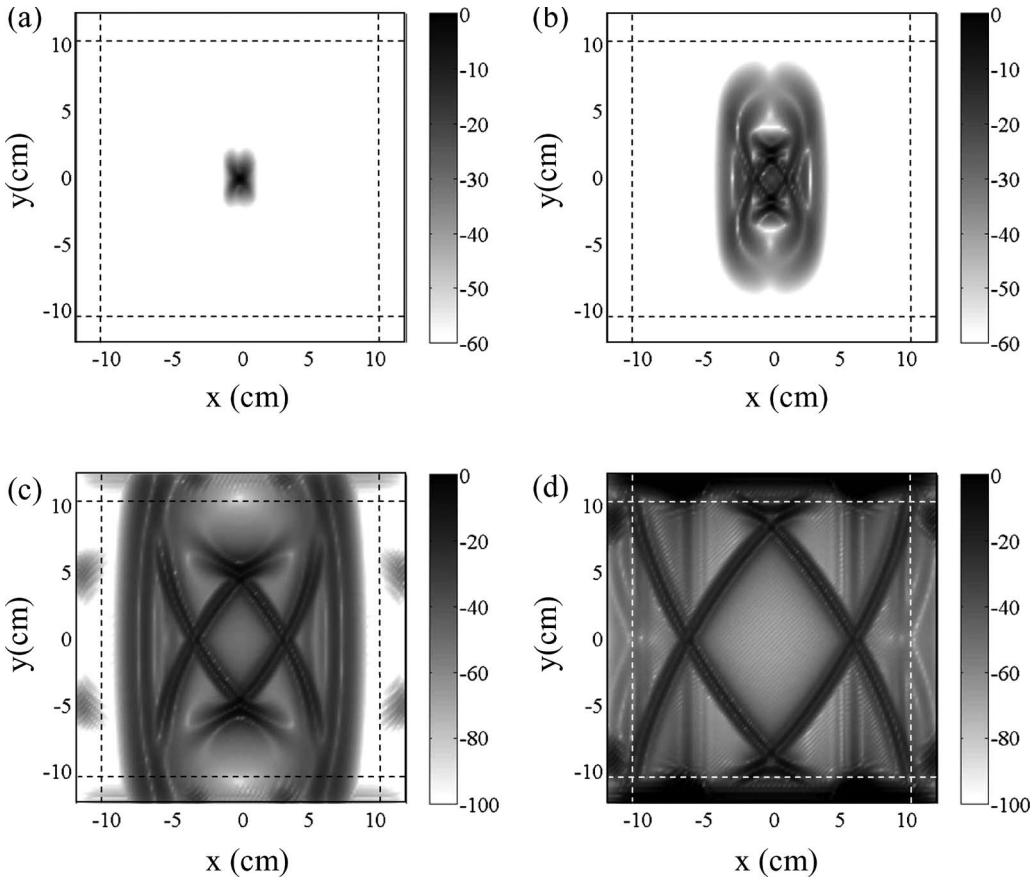


FIG. 7. Snapshots of propagation of the displacement magnitude in an orthotropic elastic medium, model III, at (a) $t=5$ μs, (b) $t=15$ μs, (c) $t=30$ μs, and (d) $t=50$ μs. The snapshots are in dB scale with a reference displacement amplitude of 10 nm. Instabilities are observed for the C-PML terminations used in the simulation.

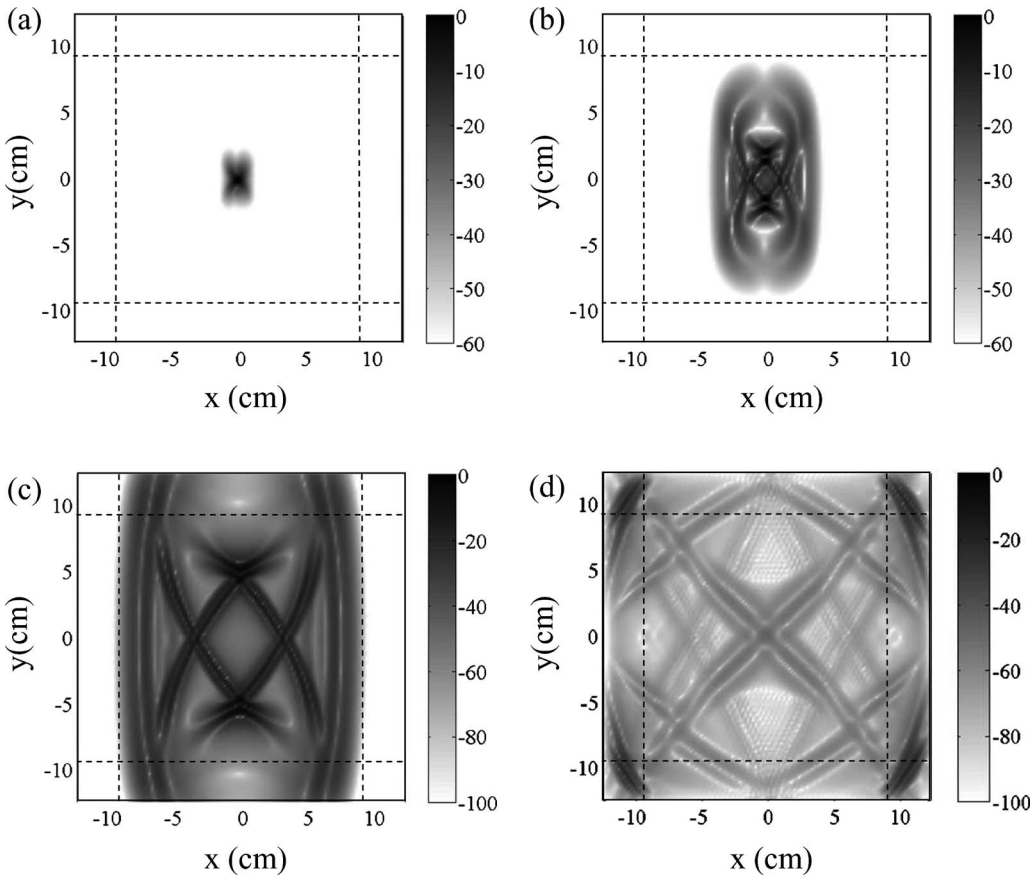


FIG. 8. Snapshots of propagation of the displacement magnitude in an orthotropic elastic medium, model III, at (a) $t=5$ μs, (b) $t=15$ μs, (c) $t=30$ μs, and (d) $t=125$ μs. The snapshots are in dB scale with a reference displacement amplitude of 10 nm. No instability is observed for the M-PML terminations used in the simulation.

of an increased length of the absorbing zone.[18,22,28] An elegant way of preserving the use of short length PML has been introduced by Skelton *et al.*,[29] but at the expense now of having one different PML for each propagating mode. For anisotropic solids an absorbing layer called multiaxial perfectly matched layer (M-PML), showing no instabilities, has recently been developed.[26] This M-PML can easily be extended to the case where the CFS stretched-coordinate metrics are used in place of the ones of PML. The introduction procedure of the M-PML is in all points similar to either PML or C-PML. However, in the M-PML, the attenuation parameters σ_x and σ_y of the stretching parameters s_x and s_y are now a function of the two space variables x and y:[18,26]

$$\sigma_x(x,y) = \sigma_{xx}(x) + \sigma_{xy}(y), \tag{23a}$$

$$\sigma_y(x,y) = \sigma_{yx}(x) + \sigma_{yy}(y). \tag{23b}$$

So, all the C-PML equations given up to now, both in frequency domain [Eq. (8)] and in time domain [Eq. (16)], can be used in the M-PML with CFS stretching, just replacing, in the expressions of s_x and s_y, $\sigma_x(x)$ by $\sigma_x(x,y)$, and $\sigma_y(y)$ by $\sigma_y(x,y)$. We define as in Ref. 26 $p^x=\sigma_{yx}(x)/\sigma_{xx}(x)$ and $p^y=\sigma_{xy}(y)/\sigma_{yy}(y)$. Results obtained for medium III with $p^x=p^y=0.25$ are shown in Fig. 8. The instabilities have completely disappeared in comparison to Fig. 7, but at the expense of a stronger reflection of the waves impinging the absorbing layer with a grazing angle [Fig. 8(d)]. More details on the behavior of these stabilized absorbing layers will be given in a future publication.

In Fig. 9, snapshots of the propagation of the amplitude of the displacement in medium V are displayed for the same four different times as in medium I. Here, contrary to the case of PML,[24] no instabilities appear even at longer time. This improved stability of the C-PML over the PML has already been demonstrated theoretically by Appelö and Kreiss.[25] In fact, the "new absorbing layer" they proposed is nothing else than a new derivation of the C-PML introduced for elastic waves in Ref. 9. It can be noted that it has also been proved that the C-PML is efficient in the case of nonlinear wave absorption.[9,30]

VI. CONCLUSION

In this paper, we extend the C-PML implementation, previously made for the first-order velocity-stress formulation, to a second-order elastic wave equation written in terms of displacements both in frequency and time domains. This new formulation has been implemented in a commercial FEM software (COMSOL MULTIPHYSICS) and in a homemade PS code. The results of the simulations demonstrate that C-PMLs have more absorbing efficiency in the cases of oblique incidence and surface wave than classical PMLs.

In the case of orthotropic material stability problems have appeared, as in the first-order velocity-stress implemen-

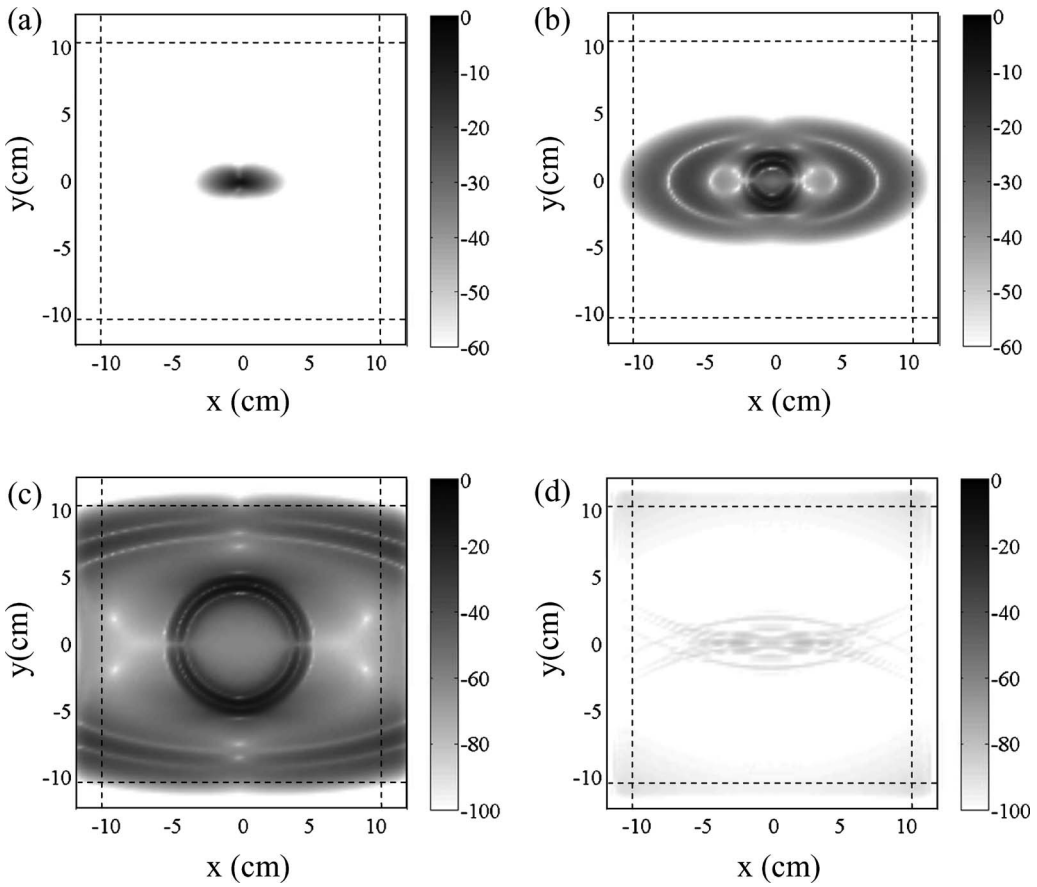


FIG. 9. Snapshots of propagation of the displacement magnitude in an orthotropic elastic medium, model V, at (a) $t=5$ μs, (b) $t=15$ μs, (c) $t=30$ μs, and (d) $t=125$ μs. The snapshots are in dB scale with a reference displacement amplitude of 10 nm.

tation. Following the "M-PML" of Meza-Fajardo and Papageorgiou,[26] we have introduced a mixture of C-PML and of sponge layer, with a controllable ratio of these two kinds of absorbing layers, in order to stabilize the C-PML, and shown that this stabilized C-PML is, in fact, not perfectly matched to the physical domain.

ACKNOWLEDGMENT

This research work was conducted, thanks to the financial support provided to the first author (Y.F.L.) in the form of a graduate scholarship by the China Scholarship Council.

[1]J. P. Bérenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[2]S. Gedney, "Perfectly matched layer absorbing boundary conditions," in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, edited by A. Taflove and S. C. Hagness (Artech House, Norwood, MA, 2005), Chap. 7, pp. 273–328.
[3]Q. H. Liu and J. Tao, "The perfectly matched layer for acoustic waves in absorptive media," J. Acoust. Soc. Am. **102**, 2072–2082 (1997).
[4]W. C. Chew and Q. H. Liu, "Perfectly matched layers for elastodynamics: A new absorbing boundary condition," J. Comput. Acoust. **4**, 341–359 (1996).
[5]F. D. Hastings, J. B. Schneider, and S. L. Broschat, "Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation," J. Acoust. Soc. Am. **100**, 3061–3069 (1996).
[6]F. Collino and C. Tsogka, "Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media," Geophysics **66**, 294–307 (2001).
[7]G. Lazzi and O. P. Gandhi, "On the optimal design of the PML absorbing boundary condition for the FDTD code," IEEE Trans. Antennas Propag. **45**, 914–917 (1997).
[8]J. A. Roden and S. D. Gedney, "Convolutional PML (CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media," Microwave Opt. Technol. Lett. **27**, 334–339 (2000).
[9]O. Bou Matar, V. Preobrazhensky, and P. Pernod, "Two-dimensional axisymmetric numerical simulation of supercritical phase conjugation of ultrasound in active solid media," J. Acoust. Soc. Am. **118**, 2880–2890 (2005).
[10]F. H. Drossaert and A. Giannopoulos, "A nonsplit complex frequency-shifted PML based on recursive integration for FDTD modeling of elastic waves," Geophysics **72**, T9–T17 (2007).
[11]F. H. Drossaert and A. Giannopoulos, "Complex frequency shifted convolution PML for FDTD modeling of elastic waves," Wave Motion **44**, 593–604 (2007).
[12]D. Komatitsch and R. Martin, "An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation," Geophysics **72**, SM155–SM167 (2007).
[13]E. Bécache, P. G. Petropoulos, and S. D. Gedney, "On the long-time behavior of unsplit perfectly matched layers," IEEE Trans. Antennas Propag. **52**, 1335–1342 (2004).
[14]D. Komatitsch and J. Tromp, "A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation," Geophys. J. Int. **154**, 146–153 (2003).
[15]W. C. Chew, J. M. Jin, and E. Michelssen, "Complex coordinate system as a generalized absorbing boundary condition," Microwave Opt. Technol. Lett. **15**, 363–369 (1997).
[16]M. Kuzuoglu and R. Mittra, "Frequency dependence of the constitutive parameters of causal perfectly matched anisotropic absorbers," IEEE Microw. Guid. Wave Lett. **6**, 447–449 (1996).
[17]Y. Zheng and X. Huang, "Anisotropic perfectly matched layers for elastic waves in Cartesian and curvilinear coordinates," in Earth Resources Laboratory 2002 Industry Consortium Meeting, Department of Earth Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA (2002).
[18]O. Bou Matar, E. Galopin, Y. F. Li, and O. Ducloux, "An optimized

convolution-perfectly matched layer (C-PML) absorbing boundary condition for the second-order elastic wave equation—Application to surface and Lamb waves propagation," in Proceedings of the 1st European COMSOL Conference, Grenoble, France (2007).
[19]S. Ballandras, D. Gachon, J. Masson, and W. Daniau, "Development of absorbing conditions for the analysis of finite dimension elastic waveguides," in Proceedings of the IEEE International Frequency Control Symposium (2007), pp. 729–732.
[20]M. Mayer, S. Zaglmayr, K. Wagner, and J. Schöberl, "Perfectly matched layer finite element simulation of parasitic acoustic wave radiation in microacoustic devices," in Proceedings of the 2007 IEEE Ultrasonics Symposium (2007), pp. 702–706.
[21]J. P. Wrenger, "Numerical reflection from FDTD-PMLs: A comparison of the split PML with the unsplit and CFS PMLs," IEEE Trans. Antennas Propag. **50**, 258–265 (2002).
[22]M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, "Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**, 1125–1133 (2004).
[23]P. Berg, F. If, P. Nielsen, and O. Skovgaard, "Analytical reference solutions," in *Modeling the Earth for Oil Exploration*, edited by K. Helbig (Pergamon, Brussels, Belgium, 1994), pp. 421–427.
[24]E. Bécache, S. Fauqueux, and P. Joly, "Stability of perfectly matched layers, group velocities and anisotropic waves," J. Comput. Phys. **188**, 399–433 (2003).
[25]D. Appelö and G. Kreiss, "A new absorbing layer for elastic waves," J. Comput. Phys. **215**, 642–660 (2006).
[26]K. C. Meza-Fajardo and A. S. Papageorgiou, "A nonconvolutional, split-field, perfectly matched layer for wave propagation in isotropic and anisotropic elastic media: Stability analysis," Bull. Seismol. Soc. Am. **98**, 1811–1836 (2008).
[27]E. Bécache and A. S. Bonnet-Ben Dhia, "PMLs for the numerical simulation of harmonic diffracted waves in an elastic plate," in Proceedings of the WCU 2003, Paris, France (2003), pp. 1019–1022.
[28]Y. F. Li, O. Bou Matar, V. Preobrazhensky, and P. Pernod, "Convolution-perfectly matched layer (C-PML) absorbing boundary condition for wave propagation in piezoelectric solid," in Proceedings of 2008 IEEE Ultrasonic Symposium (2008), pp. 1568–1571.
[29]E. A. Skelton, S. D. M. Adams, and R. V. Craster, "Guided elastic waves and perfectly matched layers," Wave Motion **44**, 573–592 (2007).
[30]D. Appelö and G. Kreiss, "Application of the perfectly matched layer to the nonlinear wave equation," Wave Motion **44**, 531–548 (2007).

Classification of a cylindrical target buried in a thin sand-water mixture using acoustic spectra

Dominique Décultot, Romain Liétard, and Gérard Maze[a)]
Laboratoire Ondes et Milieux Complexes (LOMC), FRE CNRS 3102, IUT, University of Le Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 12 March 2009; revised 8 July 2009; accepted 31 December 2009)

A number of papers have shown that it is possible to characterize an air-filled cylindrical shell immersed in water using data obtained from a backscattering spectrum. The scattered impulse time signal is constituted of echoes linked to the reradiation of waves circumnavigating around the cylindrical target. In the first part of this work, the scattered signal is calculated and then measured under conditions where the cylindrical shell is immersed in water. In the second part, the cylindrical shell is buried in a thin sand-water mixture. It is insonified perpendicularly to its axis and perpendicularly to the water mixture interface. The scattered impulse time signal is recorded and processed using a Fourier transform algorithm to obtain a resonance spectrum. Among all the resonances that are established in the explored frequency band, only those related to the circumferential S_0 wave are observed on the resonance spectrum of a cylindrical shell buried in the sand-water mixture. Resonances of the circumferential A wave also called A_0^- wave seem to have vanished. The resonance spectrum obtained by the Method of Isolation and Identification of Resonances (MIIR) reveals that it is possible to detect and classify an object buried in thin sand-water mixture in laboratory conditions.
 [DOI: 10.1121/1.3298430]

PACS number(s): 43.40.Fz, 43.20.Tb, 43.20.Fn, 43.30.Jx [ADP] Pages: 1328–1334

I. INTRODUCTION

For some years, several authors have been studying the possibility of detecting and classifying objects buried in sediment at the bottom of the sea.[1–3] In laboratory conditions, other researchers have developed methods to distinguish acoustic signals associated with buried objects from acoustic signals reverberated by the sand.[4] In this paper, the authors use the Method of Isolation and Identification of Resonances[5–9] (MIIR) to detect and to classify buried objects in a very thin sand-water mixture.

The MIIR that experimentally verifies the Resonance Scattering Theory[10] (RST) have been used in a number of papers to show that it is possible to characterize cylindrical and spherical shells using acoustic resonance spectra. The authors of the RST have demonstrated that the backscattering acoustic spectrum of a cylindrical shell, insonified by a plane wave perpendicularly to its axis, is constituted by, on the one hand, a background signal due to the reflection of the incident acoustic wave on the object and, on the other hand, by resonances associated with the propagation of circumferential waves, circumnavigating around the target in the shell or at the interface between the shell and water. The characteristics of the circumferential waves are strongly influenced by the material and the radius ratio b/a (b: inner radius; a: outer radius) of the shell.[11] In the frequency domain studied in this paper, only the resonance modes of two types of circumferential waves are observed: the A_0^- (also labeled A wave) and the S_0 wave.[12] These waves are, during their propagation, coupled with the fluid surrounding the cylindrical shell and progressively reradiate their energy into the fluid. In the case of the S_0 wave, the coupling is weak and, therefore, it can keep on propagating for several circumference turns before vanishing, whereas the A wave is significantly attenuated after only a few circumference turns (4 or 5). In this work, the influence of the sand on the propagation of circumferential waves in and around the buried cylindrical shell is closely analyzed, and resonance spectra are plotted to determine whether it is possible to detect and to classify the buried object by an acoustic method.

II. EXPERIMENTAL SETUP

The experimental arrangement is described in Fig. 1. The studied object is a finite stainless steel cylindrical shell bounded by two hemispherical endcaps.[13] Its geometrical characteristics are total length $L=70$ mm, outer radius $a=15$ mm, and the radius ratio $b/a=0.97$ (b is the inner radius). The shell wall thickness is identical on cylindrical part and on spherical part. The physical characteristics considered in the experiment are the velocity of longitudinal and shear waves, respectively, $C_L=5790$ m/s and $C_T=3100$ m/s. The stainless steel density is $\rho_{\text{target}}=7900$ kg/m^3. The object is immersed in water of density $\rho_W=1000$ kg/m^3 in which the acoustic velocity is $C_W=1470$ m/s or buried in sand-water mixture of measured density $\rho_{\text{SW}}=1250$ kg/m^3 in which the acoustic measured velocity is $C_{\text{SW}}=1650$ m/s. The degassed sand-water mixture is prepared with a clean homogeneous sand. The average dimension of sand grains is about 100 μm. The Panametrics broadband transducers (V 389) used have a central frequency equal to 500 kHz. In order to

[a)]Author to whom correspondence should be addressed. Electronic mail: gerard.maze@univ-lehavre.fr

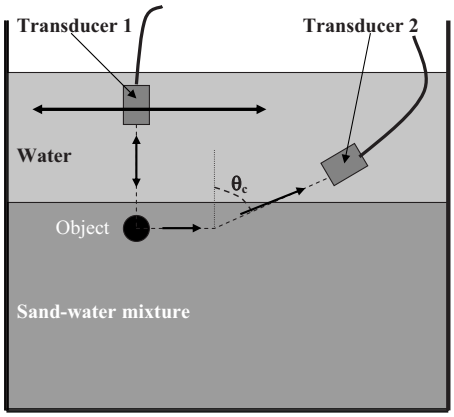


FIG. 1. Schematic of experimental setup. In monostatic method, transducer 1 is used as emitter and receiver. In bistatic method, transducer 1 is used as emitter and transducer 2 is used as receiver.

compare the acoustic attenuations in water and in the sand-water mixture, two transducers (emitter and receiver), facing each other and 30 cm apart, are successively placed in water and then in the sand-water mixture. The measured acoustic attenuation in the sand-water mixture is found to be 5 dB higher than that in water alone. In the used frequency range (100 kHz$<F<$800 kHz), the velocity and the attenuation in sand-water mixture are constant. The cylindrical object is, throughout the experiments, excited perpendicularly to its main axis.

The following experimental arrangements are realized (Fig. 1).

(i) In the monostatic method, transducer 1 working as emitter/receiver insonifies the surface of the sand-water mixture perpendicularly in the object direction.

(ii) In the bistatic method, transducer 1 is used as an emitter and transducer 2 as a receiver. A part of the energy from circumferential waves is reradiated in the sand-water mixture parallel to the sand surface, as a surface wave. The receiver detects the signal reradiated by the surface wave in water at the critical angle θ_c defined by Eq. (1):

$$\theta_c = \arcsin\left(\frac{C_w}{C_{sw}}\right) = \arcsin\left(\frac{1470}{1650}\right) = 63^\circ . \quad (1)$$

The diameter of the active surface of the transducers is 33 mm. The distance between transducers 1 or 2 and the tube is about 50 cm. This distance is larger than the beginning of the far field zone which is, for our transducers, around 37 cm. The acoustic beam diameter on the target is lightly bigger than the dimensions of the tube about 10 cm. The tube is excited by a short impulse having a large frequency band. The scattered signal is monitored and recorded by a digital oscilloscope and then transferred to a microcomputer for processing using a Fourier transform.

III. THEORETICAL RESULTS

Since the experimental cylindrical object is excited perpendicular to its axis, the associated calculated acoustic spectra are those obtained by considering an infinite length air-filled cylindrical shell immersed in a fluid. The general form of the scattered pressure field from such cylindrical shell in a plane perpendicular to the z-axis can be expressed as[5]

$$P_s(\omega) = P_0 \frac{1-i}{\sqrt{\pi k_w r}} \exp\, i(2k_w a) \exp\, i(2k_w u - \omega t) \times \sum_{n=0}^{\infty} \varepsilon_n \frac{D_n^1(\omega)}{D_n(\omega)} \cos(n\theta), \quad (2)$$

where ω is the angular frequency, $k_w = \omega / C_w$ is the wave number with respect to the wave velocity in the external fluid (C_w), P_0 is the amplitude of the incident plane wave, $D_n^1(\omega)$ and $D_n(\omega)$ are determinants obtained from the boundary conditions of the problem on the two interfaces, ε_n is the Neumann coefficient ($\varepsilon_n = 1$ if $n=0$ and $\varepsilon_n = 2$ if $n \neq 0$), u is the distance between the transducer and the surface of the cylindrical shell and r is the distance between the z-axis of the tube and the point where the acoustic pressure is calculated (the position of the emitter-receiver transducer in the case of experimental measurements). The acoustic backscattering spectrum of the studied tube obtained using Eq. (2) is presented in Fig. 2(A); a frequency window equivalent to the frequency bandwidth of the transducer is applied in the computation to simulate the sonar system response. The resonances are characterized by sharp transitions on the back-

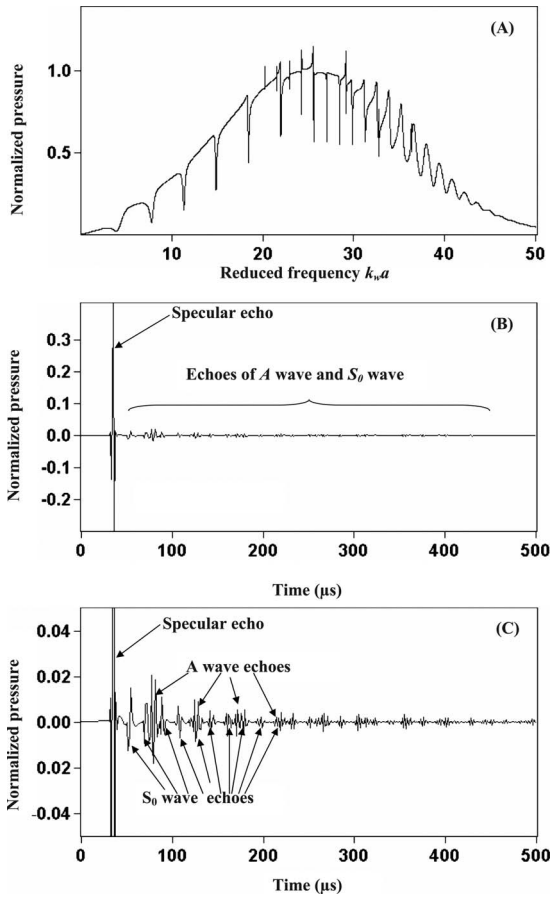


FIG. 2. Theoretical results obtained from an infinite length stainless steel cylindrical shell immersed in water and excited perpendicularly to its axis: (A) form function, (B) backscattered time signal, and (C) zoom of the backscattering time signal.

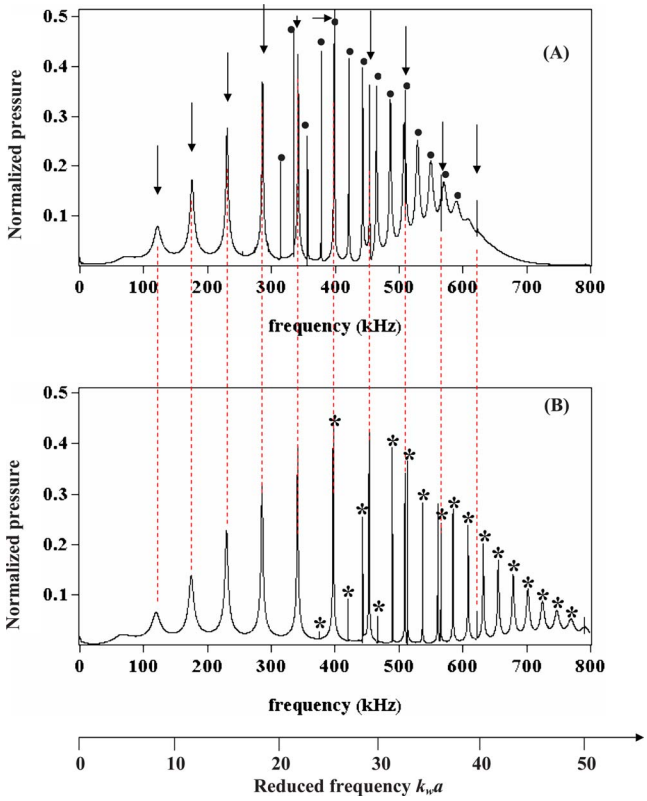


FIG. 3. (Color online) Theoretical resonance spectra. (A) Cylindrical shell in water (arrows: S_0 wave resonances, black point: A wave resonances). (B) Cylindrical shell in fluid with sand-water mixture parameters (black asterisk: A wave resonances). Dashed lines show that the resonances of the S_0 wave are at the same frequencies on the two spectra.

scattering spectrum. Resonances of the A wave are observed in the reduced frequency window $20 < k_w a < 35$. The first observed resonances of this wave are narrow (high resonance Q factor) and their width increases as the frequency rises. Resonances of the S_0 wave are wide in low frequencies and their width decreases with the increasing frequency across the studied frequency band. Figure 2(B) represents the impulse time signal backscattered from the tube, calculated applying the inverse Fourier transform on the spectrum of Fig. 2(A). On the left side of this figure, the large echo is due to the reflection of the incident wave on the cylindrical shell, it is called the specular echo. The other small echoes are associated with the A wave and the S_0 wave [Fig. 2(B)]. Figure 2(B) is plotted with a real amplitude, and it is possible to compare the amplitude of the specular echo and elastic echoes due to the energy emitted by the A wave and the S_0 wave in water. Figure 2(C) is a zoom of Fig. 2(B). The echoes of the A and S_0 waves are clearly seen. The amplitude of the S_0 echoes decreases slowly, and the coupling between this wave and the water is weak.

When the specular echo is removed and replaced with zeroes, and then the Fourier transform is applied to the time gated signal constituted of remaining elastic echoes from circumferential waves only, a resonance spectrum is obtained [Fig. 3(A)]. It is composed of peaks related to the resonance modes of the two circumferential waves. The S_0 wave resonances are indicated by arrows, whereas the A wave resonances are indicated by black points.

In the present study, this object is buried in the sand-water mixture with parameters different from those of water. Figure 3 compares the theoretical resonance spectrum of the cylindrical shell immersed in water [Fig. 3(A)] and the theoretical resonance spectrum of the same cylindrical shell immersed in a fluid with parameters of the sand-water mixture provided earlier in Fig. 3(B). The frequency scale is the absolute frequency, and the reduced frequency ($k_w a$) is calculated with the velocity of acoustic sound in water. The dashed lines show that the S_0 wave resonances are obtained at the same frequencies in the two cases. These resonances are lightly sensitive to the outside medium. This result confirms those from previous studies, indicating that the S_0 wave is mainly propagated in the shell.[14] The other resonances are due to the A wave (A_0^- wave). When the cylindrical shell is immersed in a fluid with the characteristics of the sand-water mixture ($C_{SW}=1650$ m/s, $\rho_{SW}=1250$ kg/m^3), A wave resonances are observed over a shifted frequency band ($25 < k_w a < 50$). This A wave is an interface wave, and it is sensitive to the external fluid in which the acoustic velocity is larger. In this computation, attenuation and sand particles are not taken into account.

IV. TARGET IMMERSED IN FREE WATER: EXPERIMENTAL RESULTS

Before presenting the results obtained when the cylindrical object is buried in the sand-water mixture, the experimental spectra of this target immersed in water are plotted to verify that the theoretical spectra obtained with a cylindrical shell are valid. The object is placed horizontally in a water-filled tank (diameter 3 m, height 1.5 m) and excited perpendicularly to its axis by a short impulse emitted by a broadband transducer (500 kHz). The backscattered impulse time signal from the object is detected by the same transducer operating as a receiver. A Fourier transform is applied on this time signal to obtain a backscattering spectrum [Fig. 4(A)]. In order to obtain a resonance spectrum [Fig. 4(B)], the specular echo on time signal is removed and replaced by zeroes, and a Fourier transform is applied to the filtered signal. The examination of these spectra shows a good agreement between theoretical [Figs. 2(A) and 3(A)] and experimental [Figs. 4(A) and 4(B)] results. The resonance frequencies for the two types of waves are identical. The acoustic response of the experimental object excited perpendicularly to its axis is comparable to that obtained from an infinite length cylindrical shell excited in the same conditions.[13]

V. TARGET BURIED IN THE SAND-WATER MIXTURE: EXPERIMENTAL RESULTS

In the experimental setup, the transducer beam axis is perpendicular to the surface of the sand-water mixture and to the main axis of the object (transducer 1, Fig. 1). The cylindrical shell axis is in the horizontal position and parallel to the surface of sand-water mixture. The transducer is placed at 0.5 m from this surface. The front face of the target is buried at 2 cm into the sand-water mixture. The diameter of the acoustic beam is lightly bigger than the target dimen-

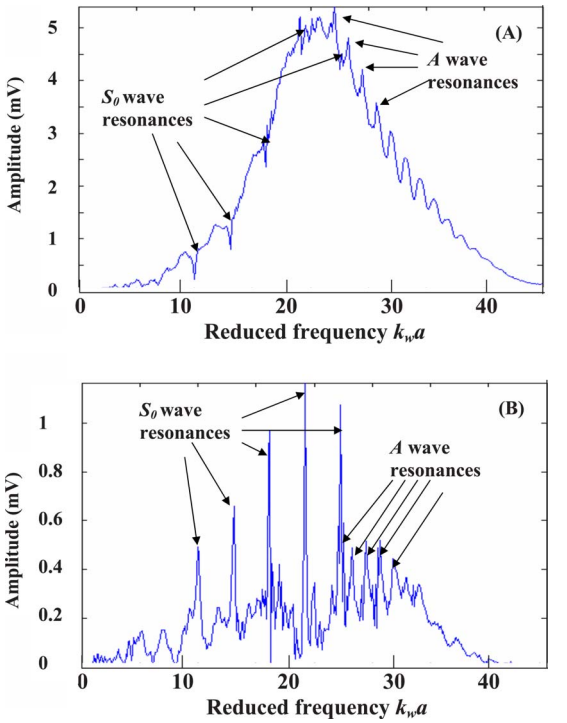


FIG. 4. (Color online) Experimental spectra obtained when the cylindrical object is in free water: (A) backscattering spectrum and (B) resonance spectrum.

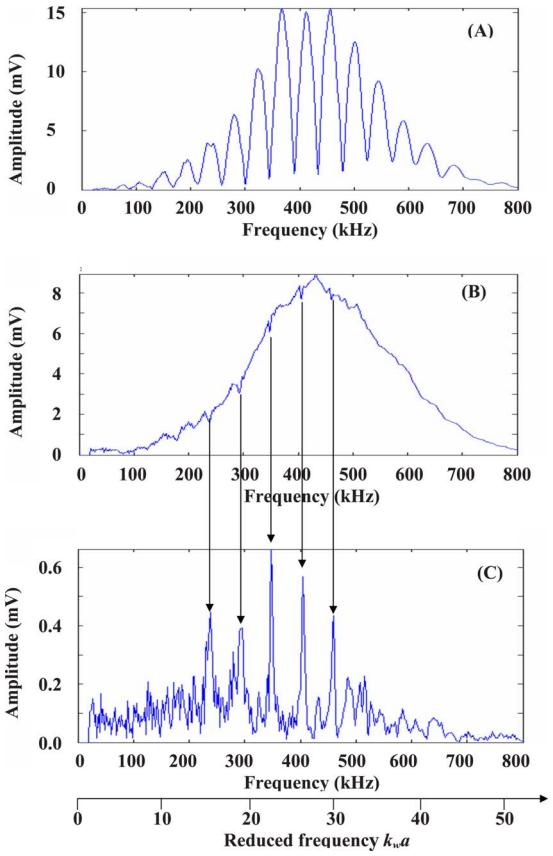


FIG. 6. (Color online) Experimental spectra obtained from the impulse time signal of Fig. 5 when the object is buried 2 cm into the sand-water mixture. (A) Backscattering spectrum obtained from the total impulse time signal. (B) Backscattering spectrum obtained after removing the reverberation echoes. Arrows indicate the S_0 wave resonance positions. (C) Experimental resonance spectra obtained after removing the specular and reverberation echoes.

sions. Figure 5 gives the backscattering time signal obtained in this condition. On left of Fig. 5(A), the first echo represents the reverberation signal, and the second major echo is the specular one. These two echoes have approximately the same amplitude and are well separated due to the burial depth. Other successive echoes are due to the propagation of the circumferential waves in and around the target [Fig. 5(B)]. Certain echoes can be due to multiple reflections between the surface of the sand-water mixture and the target surface. Figure 6(A) presents the Fourier transform of the impulse time signal from Fig. 5(A). The spectrum of Fig. 6(A) presents two differences in comparison with the spectrum obtained when the target is in free water [Fig. 4(A)].

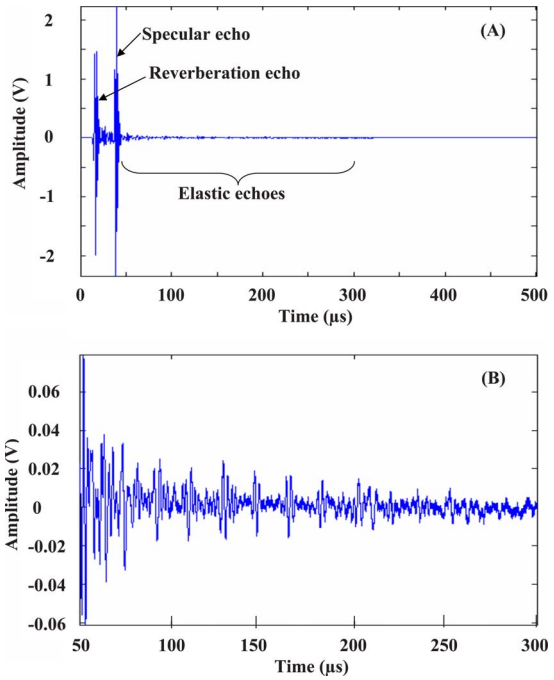


FIG. 5. (Color online) (A) Experimental impulse time signal when the object is buried 2 cm into the sand-water mixture and (B) zoom on elastic echoes.

(i) An undulation of large amplitude is observed. This undulation is due to the interference between the reverberation echo and the specular echo which have comparable amplitudes.
(ii) Resonances due to the A wave and resonances due to the S_0 wave are not visible.

The spectrum of Fig. 6(B) is obtained after having replaced the reverberation echo by zeroes in the impulse time signal in Fig. 5(A) and applying a Fourier transform on the new filtered impulse time signal. The undulation due to the interference between the reverberation echo and the specular echo is cancelled. The S_0 wave resonances are observed (arrows), whereas the influence of A wave resonances in the frequency band 25–40 has vanished.

A further time signal filtering is possible. In addition to cancel the reverberation echo from the backscattering im-

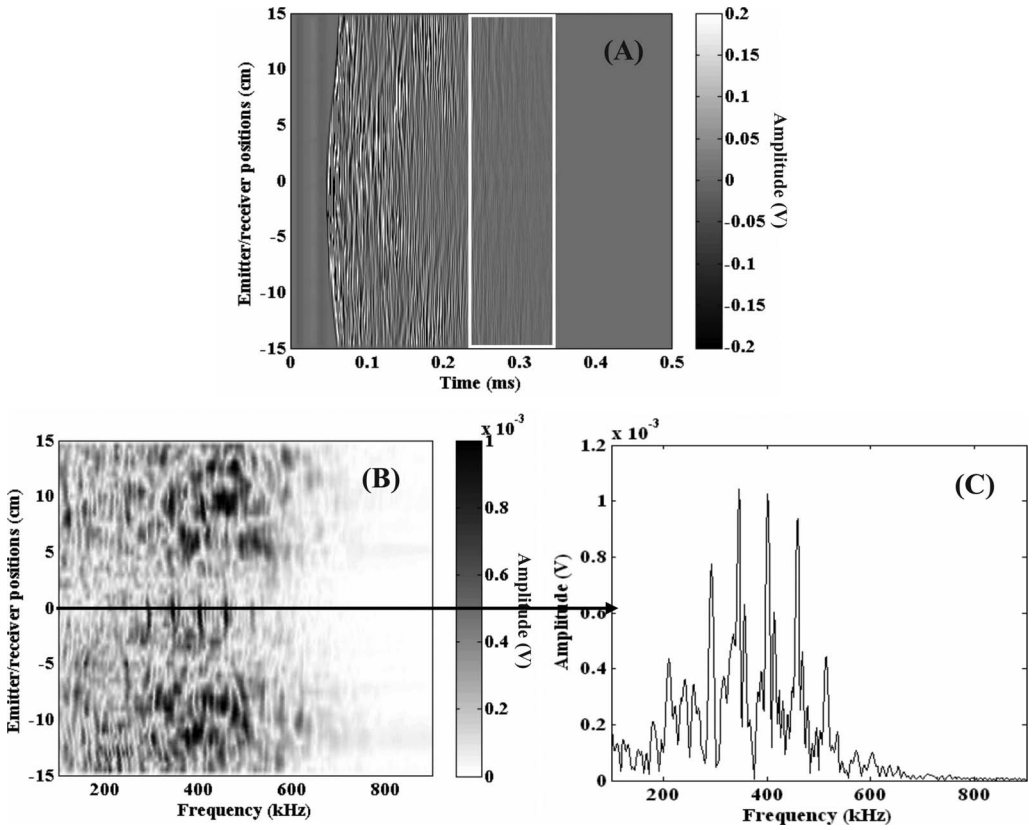


FIG. 7. Monostatic scanning of the sand-water mixture in which the object is buried (2 cm deep). (A) All the time signals, the white rectangular zone indicates the signal part used to calculate the resonance spectra. (B) All the resonance spectra. (C) A sample of resonance spectrum.

pulse time signal [Fig. 5(A)], the specular echo is also replaced by zeroes. The MIIR (Refs. 5–8) is then applied. After processing the new filtered signal using a Fourier transform, the resonance spectrum is plotted in Fig. 6(C). The background signal of the spectrum is removed, and resonances appear as isolated peaks. This spectrum shows resonances of the S_0 wave but, as seen previously, does not show resonances of the A wave. The S_0 wave is less sensitive to the outer fluid because it is a shell wave. On the contrary, the A wave which is an interface wave is strongly sensitive to the fluid surrounding the target.[12] This result is in opposition to the conclusions in Ref. 3. In this paper the echoes and the resonances of A wave (A_0^- wave) are observed for spherical and cylindrical targets buried in high-density-clutter areas of saturated sediments. To explain this difference the experimental conditions can be put forward. During its propagation around the cylindrical target, the attenuation in the sand-water mixture and the presence of the sand grains can explain the disappearance of the A wave. From the resonance spectrum, it is possible to classify the air-filled cylindrical target.

VI. SEARCHING A BURIED CYLINDRICAL OBJECT BY AN ACOUSTIC SCANNING

The experimental setup is the one presented in Fig. 1. The target is buried 2 cm into the sand-water mixture. The two following configurations are successively used.

A. Monostatic method

In the first experiment transducer 1 is used as emitter-receiver; the acoustic beam axis is perpendicular to the surface of the sand-water mixture. The transducer displacement is 30 cm around the object position (15 cm on either side from the object axis), giving the perpendicular direction to the main axis of the object. Time signals are recorded at 1 cm intervals. Figure 7(A) presents, in gray level, the assembled time signals for the explored zone (−15 cm⇔+15 cm) defined above. Echoes due to parasite reflections have been removed at the extremity of the time window. At the center of Fig. 7(A), around position 0 cm, where the cylindrical object is buried, regular echoes due to the target can be observed and, around this position a shadow effect can be seen. Figure 7(B) shows the acoustic resonance spectra calculated from the time signals surrounded by the white rectangular zone in a gray level 2D presentation. The signal outside this rectangle is replaced by zeroes. In the middle of this figure, five short black lines coming up in periodic position are observed; they are the S_0 wave resonances.

Figure 7(C), which presents the resonance spectrum at transducer position 0 cm, confirms the presence of S_0 wave resonances. As it has been mentioned before, the A wave resonances are not observed in these experimental conditions.

In Fig. 7(B) at positions ±10 cm, zones indicating black amplitudes of the same intensity as those of the resonances, due to the reverberation in the sand-water mixture,

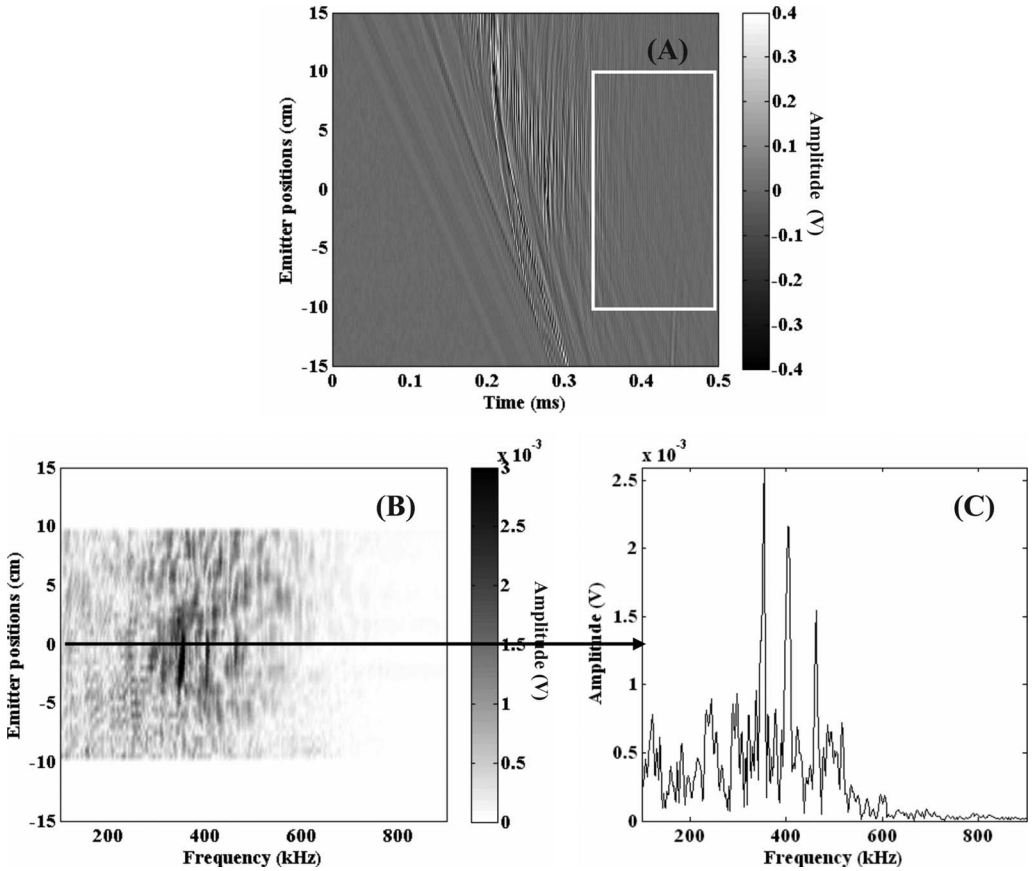


FIG. 8. Bistatic scanning of the sand-water mixture in which the object is buried (2 cm deep). (A) All the time signals, the white rectangular zone indicates the signal part used to calculate the resonance spectra. (B) All the resonance spectra. (C) A sample of resonance spectrum.

are visible. The amplitude of the reverberation signal, at these positions, is comparable to that of resonances but this signal is not periodic. At the target position, a shadow zone with a weak reverberation is observable around the periodic resonance spectra.

B. Bistatic method using surface wave propagation in the sand-water mixture

In order to complement this study, a second scanning experiment was conducted in which the incident signal is emitted by the transducer 1 (Fig. 1) perpendicularly to the sand-water mixture surface. When this incident wave excites the buried cylindrical shell, circumferential waves are generated in the shell, and part of their energy is reradiated along the sand-water mixture surface as a surface wave. Transducer 2, used as receiver, detects signals emitted by the surface wave in water at the critical angle defined earlier ($\theta_c = 63°$, Sec. II). Figure 8(A) shows the time impulse signals measured by transducer 2 for the different positions of the emitter transducer between −15 and +15 cm. Figure 8(B) is obtained after applying a Fourier transform to the signals from within the rectangular box in Fig. 8(A). This figure shows the presence of three resonance peaks associated with the S_0 wave. Figure 8(C), which is an extract from Fig. 8(B), confirms this result at emitter position 0 cm. Although the number of resonances is inferior in this bistatic method, the contrast between resonances and the reverberation signals is better and allows us to easily identify these resonances. However, this bistatic method is more difficult to realize.

VII. CONCLUSION

Experimental work presented in this paper was carried out in a laboratory tank. It shows us that resonance spectra, obtained after appropriate time signal filtering using the Method of Isolation and Identification of Resonances, can be used to classify an object buried in a sand-water mixture. To this end, two experiments were designed, one monostatic method using a single transducer working as emitter-receiver and one bistatic method using an emitter transducer and a receiver transducer detecting the reradiation of the surface wave propagating through the sand-water mixture. The resonance spectra only show resonances of the S_0 wave. This wave is very weakly sensitive to the outer medium, as it is a shell wave. Therefore, it propagates around several turns of the circumference of the tube before it vanishes. During these circumnavigations it progressively reradiates its energy. As a result, echoes due to this wave continue to be well detected after the detection of the reverberation echo and the specular one. Resonances of the A wave (A_0^- wave) are not detectable when the target is buried since the sand-water mixture strongly modifies the propagation of this interface wave.

This study has been realized to show that it is possible to identify a man-made elastic target with the resonance spec-

trum. In this paper, only the results obtained with a perpendicular excitation to the target axis are described. In a second stage, the oblique excitation to the target axis and to the sand-water mixture surface must be explored.

ACKNOWLEDGMENTS

This work was funded by DGA through Thales Underwater Systems (TUS) subcontract. The authors are grateful to Dr. Jeremiah Chiumia and to the reviewers for their careful reading of this manuscript and useful suggestions.

[1]A. Tesei, R. Lim, A. Maguer, W. L. Fox, and H. Schmidt, "Measurements of acoustic scattering from buried spherical shell," J. Acoust. Soc. Am. **112**, 1817–1830 (2002).
[2]I. Lucifredi and H. Schmidt, "Subcritical scattering from buried elastic shells," J. Acoust. Soc. Am. **120**, 3566–3583 (2006).
[3]A. Tesei, J. A. Fawcett, and R. Lim, "Physics-based detection of man-made elastic objects buried in high-density-clutter areas of saturated sediments," Appl. Acoust. **69**, 422–437 (2008).
[4]D. Décultot, K. Cacheleux, and G. Maze, "Detection and classification of an object buried in sand by an acoustic resonance spectrum method," in International Conferences on Detection and Classification of Underwater Targets, Edinburg, UK (2007), Proceeding CD Vol. **29**, Pt. 6.
[5]G. Maze, "Acoustic scattering from submerged cylinders. MIIR Im/Re: Experimental and theoretical study," J. Acoust. Soc. Am. **89**, 2559–2566 (1991).
[6]J. Ripoche, G. Maze, and J.-L. Izbicki, "A new acoustic spectroscopy: Resonance spectroscopy by the MIIR," J. Nondestruct. Eval. **5**, 69–79 (1985).
[7]J. Ripoche and G. Maze, "A new acoustic spectroscopy: The resonance scattering spectroscopy by the MIIR," in *Acoustic Resonance Scattering*, edited by H. Überall (Gordon and Breach, New York, 1992), Sec. 5, pp. 69–103.
[8]G. Maze and J. Ripoche, "New perspective on problems of classical and quantum physics," in *Experimental Acoustic Scattering*, edited by P. P. Delsanto and A. W. Saenz (Gordon and Breach, New York, 1998), pp. 157–200.
[9]P. Pareige, P. Rembert, J.-L. Izbicki, G. Maze, and J. Ripoche, "Méthode Impulsionnelle Numérisée (MIN) pour l'isolement et l'identification des résonances de tubes immergés (Numerical impulse method (NIM) for the isolation and identification of resonances of immersed tubes)," Phys. Lett. A **135**, 143–146 (1989).
[10]L. Flax, G. C. Gaunaurd, and H. Überall, "Theory of resonance scattering," *Physical Acoustics XV* (Academic, New York, 1981), pp. 191–293.
[11]G. Maze, N. Touraine, A. Baillard, D. Décultot, V. Latard, L. Derbesse, P. Pernod, and A. Merlen, "A0-wave and A-wave in cylindrical shell immersed in water: Influence on the acoustic scattering," in ASME Design Engineering Technical Conferences (DETC), Las Vegas, NV (1999).
[12]G. Maze, F. Léon, J. Ripoche, and H. Überall, "Repulsion phenomena in the phase-velocity dispersion curves of circumferential waves on elastic cylindrical shells," J. Acoust. Soc. Am. **105**, 1695–1701 (1999).
[13]N. Touraine, L. Haumesser, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, "Analysis of acoustic scattering at variable incidences from an extra thin cylindrical shell bounded by hemispherical endcaps," J. Acoust. Soc. Am. **108**, 2187–2196 (2000).
[14]J.-L. Izbicki, G. Maze, and J. Ripoche, "Influence of free modes of a acoustic scattering of a cylindrical shell," J. Acoust. Soc. Am. **80**, 1215–1219 (1986).

The effect of buildings on acoustic pulse propagation in an urban environment

Donald G. Albert
USA ERDC Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290

Lanbo Liu
Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, U-2037, Storrs, Connecticut 06269-2037

(Received 28 July 2009; revised 3 December 2009; accepted 4 December 2009)

Experimental measurements were conducted using acoustic pulse sources in a full-scale artificial village to investigate the reverberation, scattering, and diffraction produced as acoustic waves interact with buildings. These measurements show that a simple acoustic source pulse is transformed into a complex signature when propagating through this environment, and that diffraction acts as a low-pass filter on the acoustic pulse. Sensors located in non-line-of-sight (NLOS) positions usually recorded lower positive pressure maxima than sensors in line-of-sight positions. Often, the first arrival on a NLOS sensor located around a corner was not the largest arrival, as later reflection arrivals that traveled longer distances without diffraction had higher amplitudes. The waveforms are of such complexity that human listeners have difficulty identifying replays of the signatures generated by a single pulse, and the usual methods of source location based on the direction of arrivals may fail in many cases. Theoretical calculations were performed using a two-dimensional finite difference time domain (FDTD) method and compared to the measurements. The predicted peak positive pressure agreed well with the measured amplitudes for all but two sensor locations directly behind buildings, where the omission of rooftop ray paths caused the discrepancy. The FDTD method also produced good agreement with many of the measured waveform characteristics.
 [DOI: 10.1121/1.3277245]

PACS number(s): 43.40.Fz, 43.28.En, 43.28.Js, 43.20.El [RMW] Pages: 1335–1346

I. INTRODUCTION

An urban environment introduces many complicating effects on acoustic propagation that are not present in a flat open area with finite impedance ground. The presence of large, closely spaced obstacles (the buildings) induces multiple reflections, diffractions, and scattering of the acoustic waves and is the focus of this paper. Other urban effects include inhomogeneous ground conditions, for example, mixtures of grass and soil surfaces with concrete or asphalt road or parking surfaces; scattering from parked cars, other objects, and complicated facades; ground roughness and elevation changes; and different atmospheric conditions such as local wind and temperature variations caused by flow around buildings, industrial heat sources, emissions from automobiles, and differential solar loadings on different road, ground, and building surfaces. In addition, ambient noise levels and characteristics may be vastly different in urban areas compared to rural locations because of the wide variety of transportation, industrial, and cultural noise sources, and these may all vary on seasonal, diurnal, and shorter time intervals.

While many acoustic measurements have been previously conducted in real urban areas, most were measurements of traffic and other continuous noise sources, e.g., by Wyle Laboratories (1971) and Garcia and Faus (1991). Although these measurements reveal the typical urban noise environment, they provide little insight into the specifics of how acoustic waves interact with and are affected by obstacles such as buildings. Propagation measurements in cities have also been conducted for a long time, but usually these measurements used continuous sources to determine levels and reverberation times, e.g., those by Wiener *et al.* (1965), Bullen and Fricke (1976), and Iu and Li (2002), but could not reveal details about the wave interactions with buildings.

Acoustic pulses have been used recently to investigate a right-angle wall, including non-line-of-sight (NLOS) propagation (Liu and Albert, 2006). In that study, a two-dimensional finite difference time domain (FDTD) calculation was also developed and gave good agreement with the measured waveforms. Acoustic pulse measurements around a single isolated building have also been reported (Alberts *et al.*, 2008). Picaut *et al.* (2005) reported measurements using blank pistol shots in a real city street for line-of-sight (LOS) configurations. These measurements revealed the complex arrivals and extended reverberation time produced by propagation within a street canyon.

In this paper, the effect of multiple buildings on outdoor sound propagation is studied using full-scale experimental measurements that were conducted outdoors with impulsive sources so that the nature and properties of the various acoustic arrivals could be determined. The measurement methods are discussed in Sec. II, followed by examination of the results for sensors in both LOS and NLOS situations. The measurements are compared to two-dimensional finite differ-

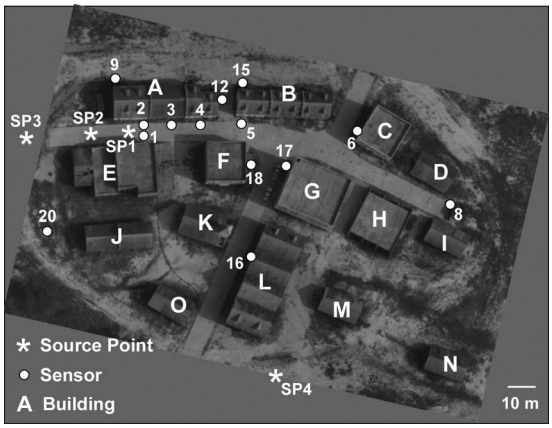


FIG. 1. Aerial photograph of the artificial training village where the measurements were conducted, showing the locations of the concrete-block buildings, shot points, and sensors. Buildings labeled D, H, I, J, M, N, and O have a single story; the other buildings were two or more stories tall. The roads are concrete, with "Main Street" running nearly horizontally in the photograph, and "Church Street" running in an approximately vertical direction. Dark ground areas are grass, and light areas are hard packed soil. The four source points are indicated by stars. Sensor numbers 7, 10, 11, 13, 14, and 19 were roof-top or interior locations and are omitted from the figure and from discussion in this paper.

ence calculations in Sec. III. The simulation results show reasonable but not exact agreement between observed and simulated time series records at 14 different sensor locations for two very different source locations. These comparisons demonstrate that the FDTD technique can be a useful tool for understanding sound propagation physics in a complex scattering environment.

II. ACOUSTIC PROPAGATION MEASUREMENTS IN AN URBAN ENVIRONMENT

A. Experimental methods

Measurements were conducted outdoors at a full-scale artificial training village composed of 15 buildings arranged over a 150×150 m^2 area. An aerial photograph (Fig. 1) shows the arrangement of the buildings along two streets; the horizontal street in the figure is referred to as Main Street, while the nearly vertical street is called Church Street. This artificial village provided a complex, realistic arrangement of full size buildings suitable for investigating one of the primary urban propagation effects: the presence of many buildings as obstacles to acoustic propagation.

The buildings were constructed of concrete blocks and most were two or three stories tall. Some of the buildings had wooden doors and strong plexiglass windows, but many had open doors and window spaces. The buildings did not have basements, but a sparse system of rain drains was located beneath the village, consisting of concrete pipes having an internal diameter of about 1 m, buried about 2 m below the ground surface.

In designing the measurements, source and sensor positions were selected to investigate both LOS and NLOS acoustic propagations, with most of the LOS measurements conducted along Main Street. These positions are also indicated in Fig. 1, and Fig. 2 shows photographic views along Main Street in the artificial village.

FIG. 2. Photographs of the artificial training village where the acoustic propagation measurements were conducted, looking in two directions along Main Street, the horizontal street in Fig. 1 (refer to Fig. 1 for the building locations). (Left) A photograph taken from the roof of building E, looking toward buildings A and B on the left side of the street. Building F is on the right side, and building C is visible in the distance at the end of the street. (Right) A photograph taken from the roof of building C looking toward building A. Building G is on the left, and building F at center left. Buildings B and C are visible on the right. Most of the line of sight measurements were conducted on this street.

Although both pressure sensors and geophones were installed at various locations within the village to record the acoustic pressure and vibration signatures produced by the explosions, only the pressure measurements at the 14 locations shown in Fig. 1 will be discussed in this paper (the other locations not discussed here included building roofs and interiors). Because of the high pressure amplitudes produced by the explosives, solid state blast pressure sensors manufactured by PCB Piezotronics (Depew, NY) were used in all locations, except at locations 8 and 20, where B&K Model 4938 microphones were used. These sensors were attached to exterior building walls at all locations, except 1, 8, and 20, and were either at 0 or 1.5 m above the ground level. Table I lists the details of the measurement installation, including the sensor models, coordinates of the source and sensor locations, and the straight-line distances between them. The actual acoustic propagation distance may be greater than that shown in Table I if the source-to-sensor path is non-line-of-sight.

The sensors were connected to two synchronized digital seismographs using multichannel geophysical cables. A Geometrics Stratavison NZ digital seismograph was used to record 48 data channels at a sampling rate of 8 kHz, while a Bison Model 9048 digital seismograph was used to record 24 additional data channels at a sampling rate of 5 kHz.

The measurements were conducted using small explosive charges consisting of 0.57 kg of C4 to produce acoustic pulses at four different source locations in the village, designated as SP1–SP4 in Fig. 1. The explosive charges were suspended from a wire at a height of 1.5 m above the ground surface and detonated using an electric blasting cap. The electrical detonation signal was also used to start the two digital seismographs. The three source positions along Main Street, spaced 15 m apart, were selected to investigate the effect of a source located in an open area or on a street with buildings on one or two sides. A fourth source location at the southern side of the village provided many complex NLOS propagation paths to the sensors on Main Street.

Table II lists the estimated minimum number of diffractions for each source-receiver pair, with LOS paths indicated by zero diffractions in the table. The source and sensor po-

TABLE I. Surveyed coordinates and straight-line distances from sensor locations to shot points.

Shot points				Distance to other shot points (m)				
Loc	*X* (m)	*Y* (m)	*Z* (m)	SP1	SP2	SP3	SP4	Description of location
SP1	0.00	0.00	0.00	0.0	15.0	29.5	103.6	Main Street shot point
SP2	−15.01	−0.02	−0.19	15.0	0.0	14.5	109.8	Main Street shot point
SP3	−29.50	0.00	−0.48	29.5	14.5	0.0	117.3	Main Street shot point
SP4	36.65	−96.87	−1.70	103.6	109.8	117.3	0.0	C-4 south of town
SP5	17.48	−57.36	−1.00	60.0	65.9	74.1	43.9	50 cal on Church Street
SP6	2.76	−4.15	5.42	7.4	19.1	33.1	99.0	50 cal on roof of bldg E

Sensors				Distance from shot points (m)					
Loc	*X* (m)	*Y* (m)	*Z* (m)	SP1	SP2	SP3	SP4	Sensor	Description of location
1	−0.01	1.03	0.00	1.0	15.0	29.5	104.6	PCB 102A	Street 1 m
2	0.06	3.82	1.50	4.1	15.6	29.9	107.2	PCB 102A04	A 4 m
3	9.24	3.70	1.50	10.1	24.6	39.0	104.3	PCB 102A06	A 10 m
4	19.69	3.28	1.50	20.0	34.9	49.3	101.6	PCB 102A06	A 20 m
5	29.80	3.12	1.50	30.0	45.0	59.4	100.3	PCB 102A09	B 30 m
6	67.67	−7.17	1.50	68.1	83.0	97.5	95.0	PCB 102A07	C 67 m
8	93.54	−28.58	1.50	97.8	112.3	126.3	88.9	B&K 4938	Street 100 m
9	−4.64	13.90	1.50	14.7	17.4	28.6	118.3	PCB 102A06	A Rear wall
12	23.50	8.42	0.00	25.0	39.4	53.7	106.1	PCB 102A06	A Around corner
15	31.89	10.34	1.50	33.6	48.1	62.3	107.4	PCB 102A07	B Rear wall
16	31.30	−42.75	1.50	53.0	63.0	74.4	54.5	PCB 106B50	L Church Street
17	44.54	−15.98	1.50	47.3	61.7	75.8	81.3	PCB 106B50	G Church Street
18	30.88	−12.92	1.50	33.5	47.7	61.8	84.2	PCB 106B50	F Church Street
20	−31.80	−30.00	0.00	43.7	34.4	30.1	95.7	B&K 4938	Ground surface

sitions provide a number of different NLOS configurations with one to six diffractions along the propagation path, providing a rich selection of complex situations to study acoustic wave interactions with the buildings.

At least five separate detonations were recorded for each source position, and the measured waveforms at all sensor locations were found to be essentially identical in appearance for the repeated shots. The peak pressure output of each explosion did vary somewhat, with standard deviation for pressures measured close to the source position of approximately 10% of the peak pressure level.

The measurements were conducted during two sunny, breezy days in November, 2001, with temperatures ranging from 8 to 19 °C, wind speeds between 2 and 5 m/s, and a

TABLE II. Estimated number of diffractions for most direct acoustic path from shot points to sensor locations.

				No. of diffractions along direct path				
Loc	*X* (m)	*Y* (m)	*Z* (m)	SP1	SP2	SP3	SP4	Description of location
1	−0.01	1.03	0.00	0	0	0	3	Street 1 m
2	0.06	3.82	1.50	0	0	0	3	A 4 m
3	9.24	3.70	1.50	0	0	0	3	A 10 m
4	19.69	3.28	1.50	0	0	0	3	A 20 m
5	29.80	3.12	1.50	0	0	0	4	B 30 m
6	67.67	−7.17	1.50	0	0	0	3	C 67 m
8	93.54	−28.58	1.50	1	1	1	3	Street 100 m
9	−4.64	13.90	1.50	2	1	1	6	A rear wall
12	23.50	8.42	0.00	1	1	1	4	A around corner
15	31.89	10.34	1.50	2	2	2	6	B rear wall
16	31.30	−42.75	1.50	2	2	2	1	L Church Street
17	44.54	−15.98	1.50	1	1	1	1	G Church Street
18	30.88	−12.92	1.50	1	1	1	2	F Church Street
20	−31.80	−30.00	0.00	1	1	0	0	Ground surface

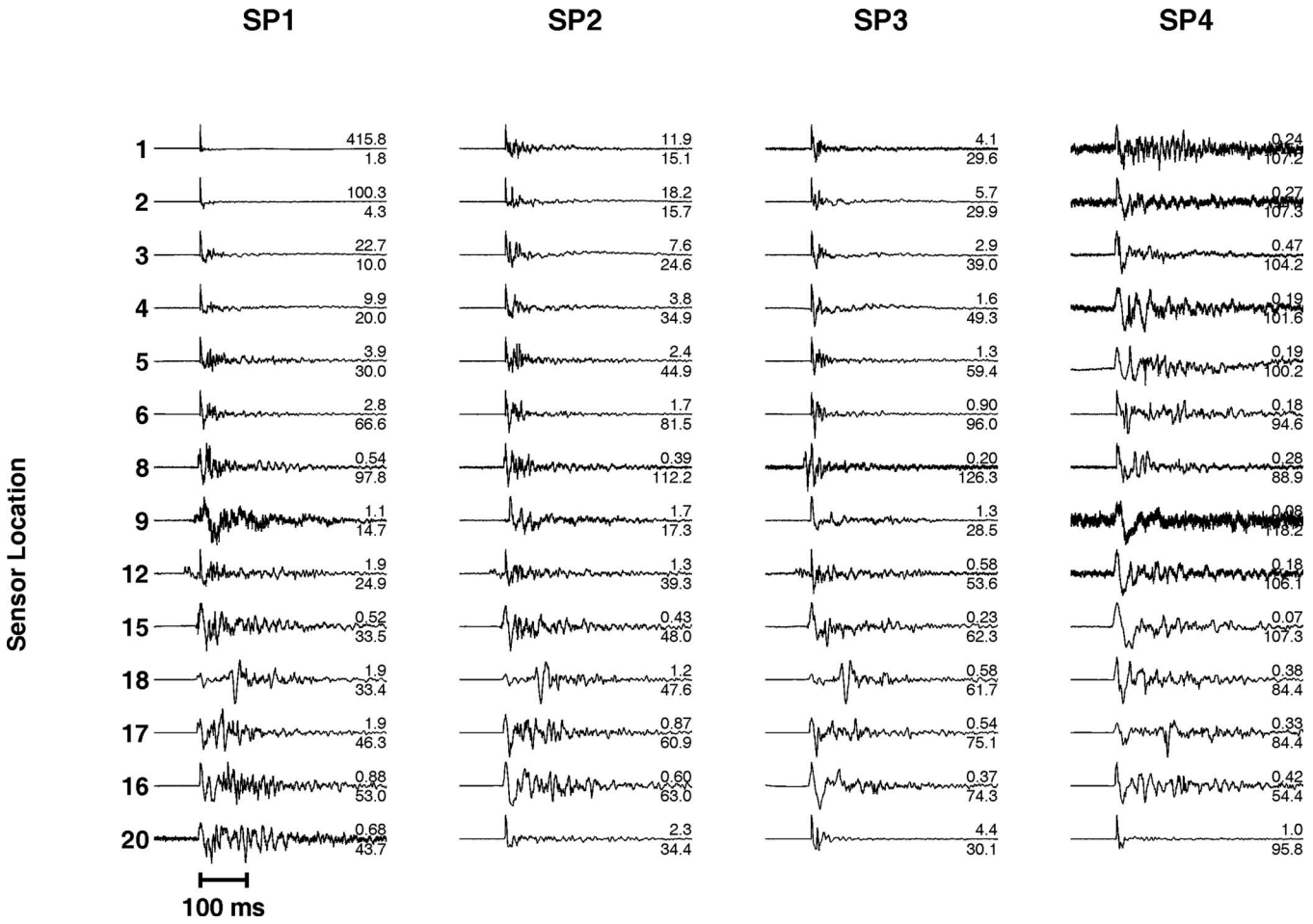


FIG. 3. Pressure waveforms recorded for all source-receiver pairs in the artificial village. The sensor number listed on the left side and the source point number at the top of each column refer to the locations in Fig. 1. The waveforms for sensors 1–6 are LOS for source points 1–3 (SP1–SP3), and the waveforms for sensor 20 are LOS for SP3 and SP4. All of the other waveforms are NLOS. The numbers at the right of each waveform are the peak pressure in kPa (top) and the straight-line source-receiver distance in m (bottom).

relative humidity of 30%–50% (Hahn and Leese, 2002). Under these weather conditions the sound wave speed varied from 336 to 348 m/s.

B. Line-of-sight propagation in a street canyon

Figure 3 shows the complete waveform data set measured for the four source positions and the 14 receiver locations. In this figure, sensors 1–6 are line-of-sight for source positions 1–3, along with sensor number 20 for source positions 3 and 4 (see Table II). The remaining source-receiver pairs are non-line-of-sight. Because of multiple reflections along Main Street, even the line-of-sight sensors show many reverberation arrivals instead of the simple pulse that is recorded for a source and receiver in an open field (a direct comparison will be presented below). The non-line-of-sight waveforms show more complexity and have longer time duration than the line-of-sight waveforms. These complex features arise from the interaction of the acoustic waves with the buildings.

All of the LOS waveforms have very short initial pulses, indicating that these arrivals are only slightly attenuated during propagation to the sensor. An examination of the LOS waveforms for SP1 shows that the visible reverberation arrivals increase as the propagation distance increases. This appearance is partially caused by the very high amplitude arrivals near the source (sensors 1 and 2) that masks the lower amplitude reverberations later in time at the plotting scale.

The measured peak positive pressure for LOS propagation, as a function of distance, is shown in Fig. 4. These measurements show that the pressure levels are slightly higher than those predicted by the ANSI Standard (ANSI, 1983) due to the hard concrete surfaces present for the measurements compared to the normal soil surface condition assumed in the predictions.

The least-squares method was applied to the measurements to determine the coefficients in the equation

$$P(r) = P_0 r^{-\alpha}, \tag{1}$$

where $P(r)$ is the peak pressure (Pa) at propagation distance r (m), and α is the attenuation coefficient. An attenuation value of 1.52 was determined, which is a high attenuation rate compared to a value of 1.12 for ANSI S2-20 predictions (ANSI, 1983) for open areas far from the source. This higher rate is likely caused by the strong amplitude decay that occurs in the high-amplitude, nonlinear propagation region close to the source.

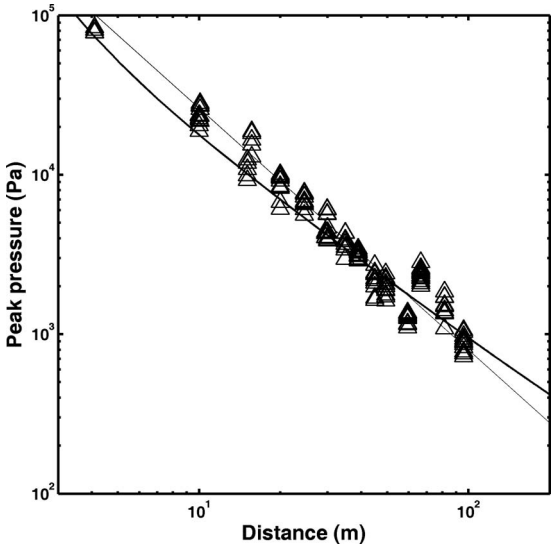


FIG. 4. Peak pressure as a function distance in an urban environment for LOS sensors. The heavy line is the ANSI Standard S2-20 prediction, while the light line is a least-squares fit to the measured data. This line gives a decay rate of $R^{-1.52}$. The sensors at 66 m (sensor location 6 in Fig. 1) were mounted on a wall nearly perpendicular to the incoming wave and show the effects of "pressure doubling" from the incident and reflected waves at the wall.

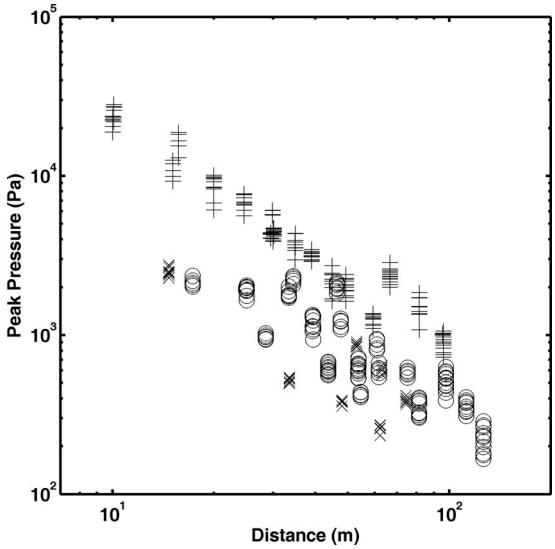


FIG. 6. Measured peak pressure as a function of distance for LOS (plus-signs) and NLOS sensors that have diffracted around one building corner (open circles) or two corners (X-signs).

Because the explosions that were used as the source of the acoustic pulses produced very high amplitudes (over 400 kPa at 1 m), nonlinear effects caused the propagation speed to depend on the pulse amplitude and to decrease with distance from the source as the pulse amplitude decreased. Figure 5 shows the acoustic velocity measured for the LOS propagation down Main Street. The wave speed rapidly decreased from around 420 to 390 m s^{-1} at a pressure level of about 10–20 kPa (at about 10 m distance from the explosion), and continued to decrease, reaching 365 m s^{-1} at 4 kPa (30 m), 360 m s^{-1} at 2.5 kPa (40 m), and 350 m s^{-1} at 1 kPa (90 m). The wave speed is also expected to decrease as a strong pulse diffracts into a shadow zone, because of the amplitude reduction caused by the diffraction itself, leading to propagation kinematics that are more complicated than a simple constant acoustic velocity situation. This variation in wave velocity can cause errors on the order of 10–30 ms (a few meters) if a constant acoustic wave speed is used for calculations or comparisons of the data with linear acoustic theory. This factor will be discussed further in the theoretical waveform comparisons presented below.

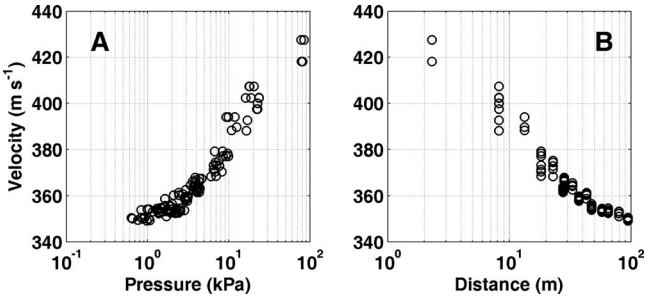


FIG. 5. (a) Measured acoustic wave velocity as a function of peak pressure for LOS paths in the village. The dependence of the velocity on amplitude is a well known nonlinear acoustic effect. (b) Measured acoustic velocity as a function of propagation distance along Main Street.

C. Non-line-of-sight propagation in an urban environment

The NLOS waveforms shown in Fig. 3 are generally more complex, with many more visible wave arrivals, and with longer time durations than the LOS waveforms recorded at similar propagation distances. In addition, the initial arrival pulse widths tend to be longer, and the peak amplitudes (indicated to the right of each trace) are lower for the NLOS than for the LOS waveforms, as shown in Fig. 6. This figure shows the measured peak pressure amplitudes for LOS and NLOS propagation paths. Here, it is difficult to assign a distance to the NLOS data because many path lengths are involved (and as shown above, the shortest path is sometimes not the peak amplitude path), so the data have been plotted using the straight-line distance as if no buildings were present. If the correct path length were used, all of the NLOS data points would shift to the right (to longer propagation distances) on the plot. The graph shows that the NLOS measurements are always smaller than the LOS measurements for identical distances, a result of the reduction in amplitude when the waves diffract into a shadow zone and also of the larger geometrical spreading from the longer path lengths involved in multiple reflections. In addition, the plot shows that multiply diffracted waves generally have smaller amplitudes than those waves that were only diffracted once.

Two exceptions to the above observations should be noted. First, the amplitude for the single-diffraction NLOS SP1-Sensor 17 path (approximately 2 kPa at 46.4 m) is nearly the same as for LOS positions at similar distances (these measurements are the circles lying surrounded by plus-signs at about 50 m distance in Fig. 6). The relatively small "extra NLOS" attenuation is reasonable since this sen-

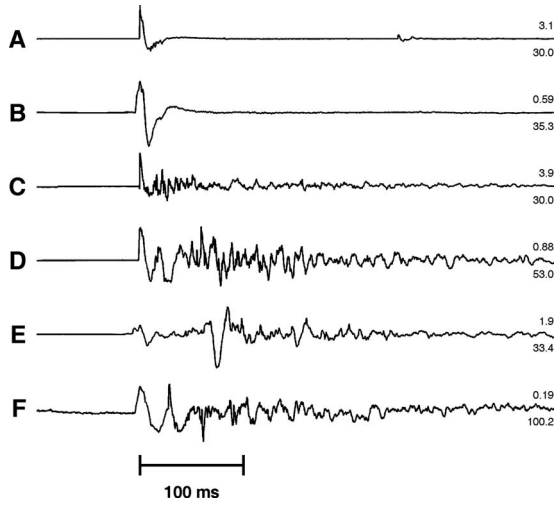


FIG. 7. Pressure waveforms produced by detonation of C4 explosive 1.5 m above the ground and recorded by pressure sensors also located 1.5 m above the ground for various geometries. The top number on the left of each trace is the peak pressure in kPa, and the lower number is the straight-line distance from the source in m. The waveforms have been time-shifted to align the first arrivals. (a) Line-of-sight propagation in open air above grass, 30 m distance. (b) Propagation around a single corner, 35 m distance (from Liu and Albert, 2006). (c) Line-of-sight propagation down a street with buildings present, 30 m. (d) Propagation around a corner with buildings present, 53 m straight-line distance (through buildings). (e) Propagation around a corner with buildings present, 33 m straight-line distance (through a building). (f) Propagation around two large buildings, 100 m straight-line distance (through two buildings). Trace B was produced using an explosive charge size of 0.28 kg, while all of the others were produced using 0.57 kg. Waveforms C, D, and E were recorded from the same detonation, but at different sensor positions (16, 18, and 5) in the urban village.

sor, located on building G, is barely within the diffraction shadow zone of building F (approximately 22°), so diffraction losses are small.

The second anomaly is that the SP1-Sensor 16 path (the X-signs above the circles in Fig. 6, with amplitude approximately 875 Pa at 53.0 m) involves two diffractions around buildings E and K (with diffraction angles of about 14° and 6°, respectively), but has a higher amplitude than the singly diffracted measurements at nearly the same distance (SP3-12, about 580 Pa at 53.6 m). In the latter case, the peak arrival is a later reflection from building B (and so travels a longer path) that has undergone a relatively large angle diffraction of about 25° around building A, thus producing a greater reduction in amplitude. The combination of these path factors is apparently enough to produce a lower amplitude for this NLOS arrival, compared to others at similar distances.

Figure 7 provides a detailed view of these properties by comparing a few typical waveforms recorded in the urban environment with those recorded in an open field or near a single, right-angle wall (Liu and Albert, 2006). The parameters of the waveforms shown in Fig. 7 are listed in Table III. Figure 7(a) shows a waveform recorded in open air over grassland, while Fig. 7(c) shows the waveform recorded for the same propagation distance, but for LOS propagation down Main Street in the village. In both cases, the propagation distance was 30 m. These measurements show that in the urban environment, the waveform contains higher frequencies (the initial part of the waveform is shorter in time), has a higher amplitude (peak positive pressure 3.9 kPa vs 3.0 kPa) probably because of the higher ground impedance of the concrete compared to the soil surface, and is more complex than the open field waveform. The many arrivals visible after the first main arrival in the urban waveform are from the multiple reflections from the buildings in the street canyon. These reflected arrivals are absent for the waveform recorded in the open field. These differences between open field and urban waveforms were observed at many different propagation ranges, from a few meters to 100 m, as long as the acoustic propagation remained line-of-sight (Fig. 3).

Waveform B in Fig. 7 was recorded over a similar distance (35.3 m) after diffraction around a single corner, displaying the low-pass filtering effect of the diffraction (Pierce, 1989; Liu and Albert, 2006). The diffraction-induced broadening of the waveform is very noticeable and can be used to identify likely diffraction arrivals in the more complex NLOS traces, such as those shown labeled D, E, and F in Fig. 7. These three NLOS urban waveforms also show a characteristic found in many NLOS situations, that the peak pressure is sometimes a later arrival rather than the first arrival as for LOS and other simpler geometries.

Referring to Fig. 1 for the locations, the waveform D shown in Fig. 7 was produced by propagation from source point 1 to sensor 16 on building L, a straight-line distance of 53 m. The most direct propagation path has diffractions around two buildings: E at about 14° and K at about 6°. The first arrival shows a waveform with similar duration as the simple diffraction waveform B, and the low frequency arriv-

TABLE III. Characteristics of the waveforms shown in Fig. 7.

		Positive peak		First arrival (if different from peak)			
	Distance (m)	Pressure (kPa)	Arrival time (ms)	Pressure (kPa)	Arrival time (ms)		
A	30.0	3.05	78.8			Open field	(Liu and Albert, 2006)
B	35.3	0.55	108.0			Around one corner	(Liu and Albert, 2006)
C	30.0	3.91	79.9			Urban LOS	SP1, sensor 5
D	53.0	0.858	209.2	0.877	150.0	Urban NLOS	SP1, sensor 16
E	33.4	1.55	197.2	0.516	112.6	Urban NLOS	SP1, sensor 18
F	100.2	0.138	321.9	0.126	293.4	Urban NLOS	SP4, sensor 5

Refer to Fig. 1 for the source point and sensor locations for the urban waveforms.

als are tentatively identified as probable diffractions. The second arrival at 19 ms later follows the same path after reflecting from building B; this identification is based on the time difference corresponding to a path length approximately 7 m longer and on theoretical calculations discussed below. Thus, waveform D is very complex, showing at least two diffracted arrivals and many multiple arrivals.

Waveform E in Fig. 7 was recorded for the same explosion as for waveform C, but with the sensor located around a corner, shielded by building F. The straight-line distance between the source and receiver, ignoring the presence of the building, was 33 m. This is a very complex waveform, with some lower frequency arrivals visible. The first arrival can be identified as a simple diffraction around the Main Street corner of building F (the upper right corner in Fig. 1); the second low frequency arrival (at approximately 180 ms) is the reflection from building G after slightly diffracting around the same corner of building F; and a third low frequency arrival is the ray path that has reflected from G to F to G and back to F (at 265 ms). The second arrival has the largest amplitude despite the longer propagation distance because the diffraction angle is only slightly into the shadow zone compared to the first arrival that diffracted around an angle of 90° deep into the shadow.

Waveform F is for a longer propagation distance of 100 m for the path from source point 4 to sensor 5. In this case, there are two very large buildings directly in the straight-line path between source and receiver, and many surrounding buildings present. Again, the peak amplitude occurs after the first arrival, and there are low frequency portions in the waveform that are likely produced by diffractions, along with higher frequency arrivals from multiple reflections.

To summarize the results of this section, Figs. 6 and 7 show the complexity involved in NLOS propagation. Wave arrival times, as a function of distance, are not straight lines as for propagation in an open field because of the many different propagation paths arriving at a sensor, including longer diffraction and multiple reflection paths. In addition, the waveforms displayed in Fig. 7 show some of the complexity that a sensor or human listener will encounter when attempting to identify and locate sources in urban areas.

III. ACOUSTIC PROPAGATION MODELING IN AN URBAN AREA

The measurements discussed in Sec. II show that a simple acoustic source pulse produces a very complex pressure time series when recorded after propagation in an urban location because of the multiple interactions with the buildings. In this section, a computational model of the acoustic propagation that includes these effects is demonstrated. The interest in this paper is in what Kang (2006) called *microscopic* urban modeling, since the intent is to predict sound propagation for relatively short propagation distances of tens to hundreds of meters, compared to less-detailed *macroscopic* models intended to produce city-scale maps of noise levels.

Although many methods for predicting sound levels in urban areas have been developed over a long period of time (Lee and Davies, 1975; Oldham and Radwan, 1994; Heutschi, 1995), theoretical predictions are still a topic of research interest. Some recent studies have used statistical methods to measure or model traffic noise, or diffusion approaches to predict noise levels in street canyons with complex building facades (Brown and Lam, 1987; Garcia and Faus, 1991; Picaut *et al.*, 1999; Kang, 2000, 2001). Kang (2006) provided a thorough review of past work on urban sound predictions. A promising approach to calculate detailed time series predictions is the FDTD method (Botteldooren, 1994; Blumrich and Heimann, 2002). This method can handle complex geometries and acoustic pulses, but is limited to some extent by computer time and memory requirements. The method has been previously used to study urban acoustic propagation by a number of authors (Albert *et al.*, 2005; Liu and Albert, 2006; Van Renterghem *et al.*, 2006).

A. FDTD propagation model

A FDTD acoustic propagation model is applied in this paper since it can include all of the waves and interactions with buildings produced by an impulsive point source.

The FDTD acoustic propagation model used in this paper has been discussed in a previous publication (Liu and Albert, 2006), so only a brief summary of the method is given here. This finite difference model is based on the expression of acoustic propagation as a set of first-order, velocity-pressure coupled differential equations, similar to the motion and continuity expressions used by other authors (Virieux, 1984; Wang, 1996). To approximate the derivatives in the acoustic wave equation with finite differences, a staggered difference algorithm proposed by Yee (1966) is used in a two-dimensional spatial domain. The marching is also staggered between the computations of the air pressure and the particle velocity in the time domain. The perfectly matched layer technique (Berenger, 1994) was adapted for the absorption boundary condition and achieved highly effective suppression of reflections from the domain boundaries.

The computational effort is reduced by restricting the calculations to a simplified two-dimensional representation of the urban area, and this simplification allows the calculations to be made on a desktop personal computer. Buildings are treated as solid concrete blocks and a source function representing an explosion waveform with a peak frequency of about 50 Hz (Liu and Albert, 2006) was used for the calculations. This power spectral density of this source function is 1/30 of the peak level at 200 Hz, limiting the calculations to this relatively low frequency bandwidth.

For many finite difference calculations the Courant stability criterion (Wang, 1996) (nominally ten time sample points per wavelength) ensures that the computations will be stable and accurate. However, a more stringent criterion (Schroder and Scott, 2002) was needed for these calculations because of the very high impedance contrast between concrete, with a wave velocity of 2950 $\mathrm{m\,s^{-1}}$ and density 2300 $\mathrm{kg\,m^{-3}}$, and air, with a speed of about 350 $\mathrm{m\,s^{-1}}$ and density of 1.2 $\mathrm{kg\,m^{-3}}$. The computations were made stable by lowering the time and spatial step sizes in the computation and by reducing the impedance contrast between the materials.

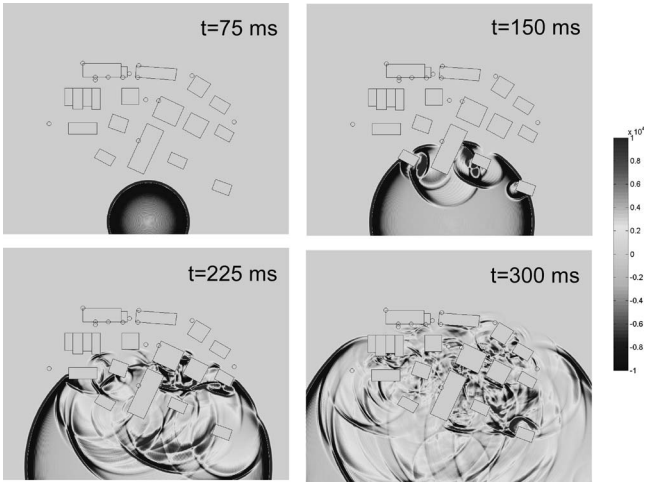


FIG. 8. Calculated acoustic propagation within the artificial village for a source located at SP4. Snapshots of the simulated wavefield are shown for times $t=75$, 150, 225, and 300 ms.

The impedance contrast was reduced by using a lower value of 23 kg m^{-3} instead of the true value of 2300 kg m^{-3} for the density of concrete. This change preserves the wave velocities and the kinematic results while still giving sufficient accuracy to the reflection amplitudes, as shown by Liu and Albert (2006). A spatial grid of 0.1 m was selected along with a time step of 0.01 ms. These choices not only improve stability, but also increase the accuracy of the calculations for buildings that are skewed with respect to the Cartesian grid coordinates. The village was thus represented by a 2000 $\times$1400 spatial grid: a total of 2.8 million grid points. The computations were performed using the MATLAB language.

Corrections for the two-dimensional geometric spreading (an additional factor of $r^{-1/2}$, where r is the propagation distance) and the effect of the ground surface (an additional factor of 2, assuming a rigid ground surface and that the direct and reflected path lengths are nearly identical) were applied in the calculations. With these corrections, the model yields accurate results for simple LOS and NLOS geometries (Liu and Albert, 2006). Because of the two-dimensional approximation, acoustic energy outside the plane of the propagation model is ignored; i.e., propagation over the top of the buildings or diffraction from upper edges is not included in the calculations.

B. Simulation of acoustic pulse propagation in a village

Movies and snapshots of the computed wavefields are some of the very useful output products produced by the FDTD computations; these displays allow the complex wave interactions with the buildings to be studied in detail. Since snapshots for a source located at SP1 have already been published (Albert *et al.*, 2005), the wavefield snapshots for a explosion source located at SP4 are shown here in Fig. 8. At $t=150$ ms, the initial pulse has entered the village, and reflections and diffractions around the buildings are clearly visible and identifiable. As the propagation time increases, the wavefield within the village reflects and scatters around the buildings producing very complex interference patterns. Urban propagation features, including the diffraction from building corners and narrow gaps between buildings, the canyon effect along a street, and the pressure shadow zones behind buildings, are visible in these snapshots.

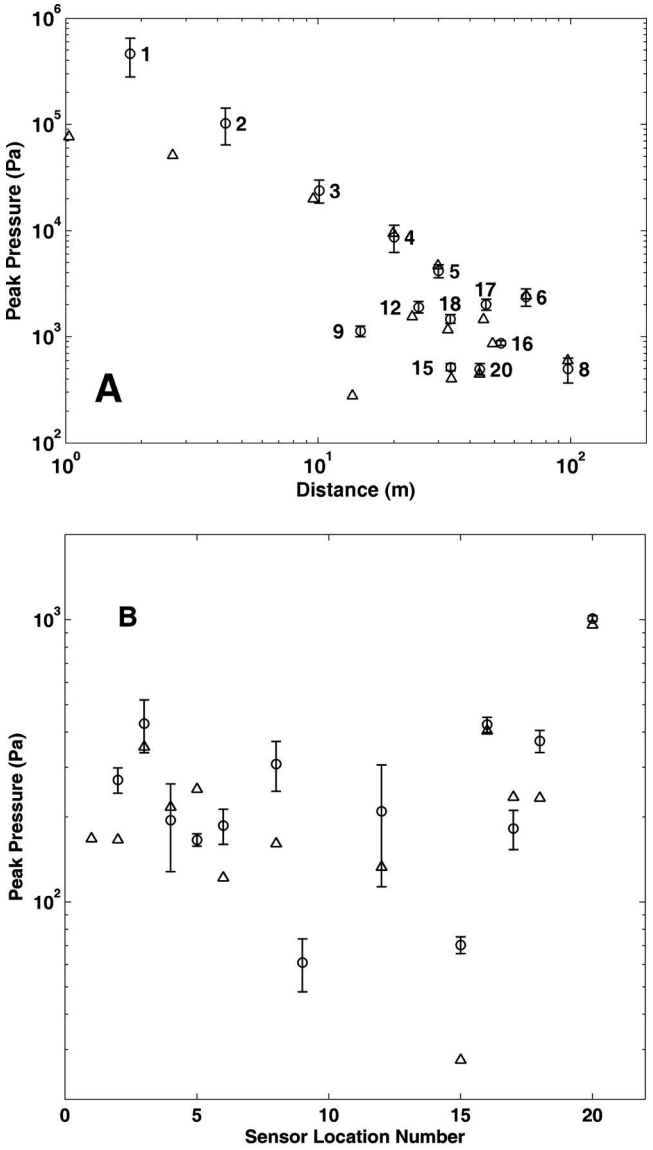


FIG. 9. (a) Comparison of measured and calculated peak pressure amplitudes as a function of propagation distance for SP1. Circles represent the mean measured peak amplitudes with error bars of two standard deviations. Triangles are the peak amplitudes calculated using the two-dimensional FDTD method. NLOS sensors (with lower amplitudes) are plotted at the LOS distance for that sensor location. Numbers near the measurements indicate the sensor location in Fig. 1. (b) Comparison of measured and calculated peak pressure amplitudes for each sensor location (see Fig. 1) for SP4. Circles represent the mean measured peak amplitudes with error bars of two standard deviations. Triangles are the peak amplitudes calculated using the two-dimensional FDTD method. All sensors except number 20 are NLOS. The calculated value for sensor No. 9 (behind building A) is 1.2 Pa, below the lower boundary of the plot.

A comparison of the measured and calculated peak amplitudes for a source located at SP1 is shown in Fig. 9(a). The initial source pulse amplitude used for the calculations was estimated as 77 kPa at 1 m, based on the measured values for the Main Street sensors at 10–30 m distance; this source amplitude was subsequently used in all of the theoretical calculations for all source positions. Figure 9(a)

shows good agreement between the measured and calculated amplitudes, except for a few locations. The largest disagreement is for the two closest sensor positions (1 and 2) along Main Street, where the calculated pressure amplitudes are much lower than the measured amplitudes. This disagreement is expected because the nonlinear pressure effects in the measurements at these sensor positions are not included in the linear acoustic calculations. The next largest disagreement occurs for the sensor directly behind building A [plotted at 12 m distance in Fig. 9(a)], where the calculated amplitude is lower than the measured amplitude by about a factor of 3. At this sensor location, the two-dimensional computational model requires the waves, either to diffract around two right-angle exterior walls or to transmit through two concrete-block walls. The path over the roof was not included in the simulation, and it will be shown below that this path has a significant effect on the measured amplitude. In addition, there was probably a much higher transmission coefficient directly through the building because there were many large open windows and doors in building A that would allow the high amplitude source pulse to penetrate into the building virtually without attenuation. In the theoretical calculations, very little acoustic energy can enter the building because of the high impedance contrast between air and the concrete building walls. For all of the other NLOS sensors, the calculated amplitudes are close to the measured amplitude range. It is noteworthy that the measured amplitude for sensor 15, behind building B, is much lower than for the sensor behind building A and that the agreement with the calculated amplitude at that location is good.

The measured and calculated pressure amplitudes for SP4 are shown in Fig. 9(b). In this case, most of the sensors are about 100 m from the source, so the results are plotted by sensor position number (Fig. 1) instead of by propagation distance. The same theoretical source amplitude of 77 kPa at 1 m distance was used for these calculations. Five of the sensor amplitudes agree within the experimental limits, six give lower theoretical amplitudes, and two give higher theoretical amplitudes. All but one of these cases (number 20) is NLOS with multiple building interactions, and the amplitude agreement is generally within better than a factor of 2, with two exceptions: sensors 9 and 15, which were mounted on walls directly behind buildings A and B, respectively. As mentioned above, computational results for these sensors are likely to be less accurate because of the neglect of propagation paths over the tops of buildings lowering the calculated levels and the actual propagation paths through the open doors and windows into the buildings increasing the measured levels.

Waveform modeling is a very stringent test of any computational propagation model, since it requires accurate amplitude and phase agreement with measurements over a broad band of frequencies. A comparison of the measured pressure waveforms with those calculated using the finite difference method for SP1 is given in Fig. 10. For these calculations, a time-varying velocity for acoustic waves in air, starting at 380 m s^{-1} and reducing to 350 m s^{-1} after 300 ms, was used to approximate the measured pressure-dependent sound speed shown in Fig. 5. In Fig. 10, the upper

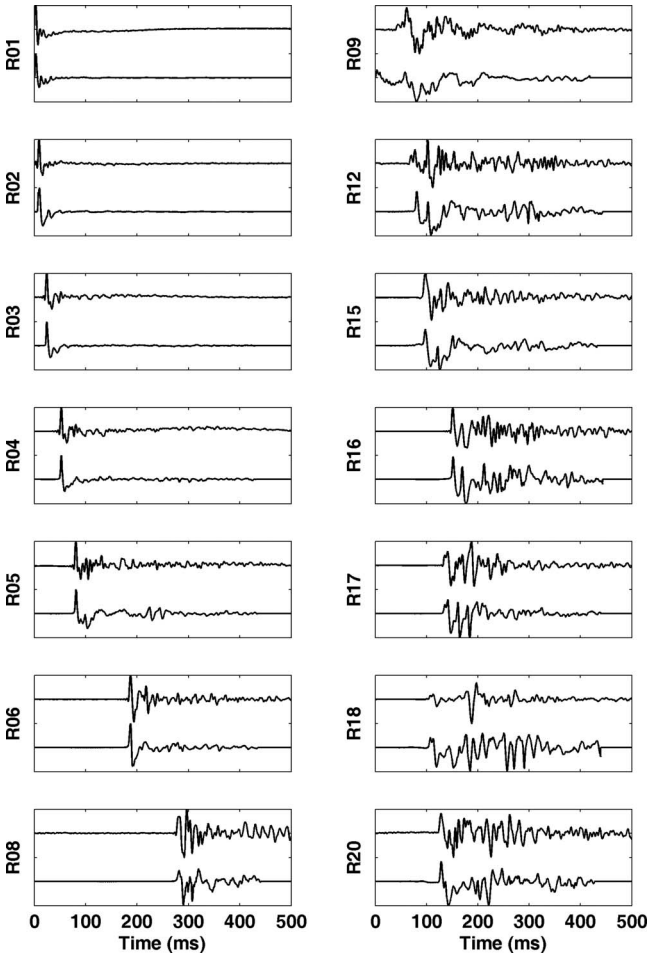


FIG. 10. Measured (upper) and theoretical (lower) waveforms at 14 locations in the artificial village for source point 1. The observed waveforms have been low-pass filtered at 200 Hz, and the calculated waveforms have been time-shifted to correct for variable acoustic velocity as explained in the text. These waveforms are all individually normalized. The side labels give the locations, as shown in Fig. 1.

traces in each panel are the measured waveforms, while the lower traces are the calculated time series for each sensor location. The measured waveforms were low-pass filtered with a passband below 200 Hz to adjust their frequency content more closely to the calculated frequency band. The difference in frequency content was expected because the calculations were restricted to low frequencies by the choice of the theoretical source function and for stability considerations, as discussed earlier.

In addition, a time shift was applied to each of the calculated waveforms to align the calculated first arrivals with the measured time series. The largest time shift applied was 12.6 ms for sensor 16, while two other NLOS sensors were shifted by 12.3 and 10.1 ms, and the remaining 11 sensors were shifted by less than 10 ms. These time shifts were needed because a single velocity function cannot account for all of the propagation paths, since the peak pressure (and thus the wave velocity) will decrease rapidly once a diffraction corner is turned, so the wave velocity may be different for different parts of the wavefield at the same time instant. In fact, the velocity for different arrivals within the same time series is often different because of the presence or ab-

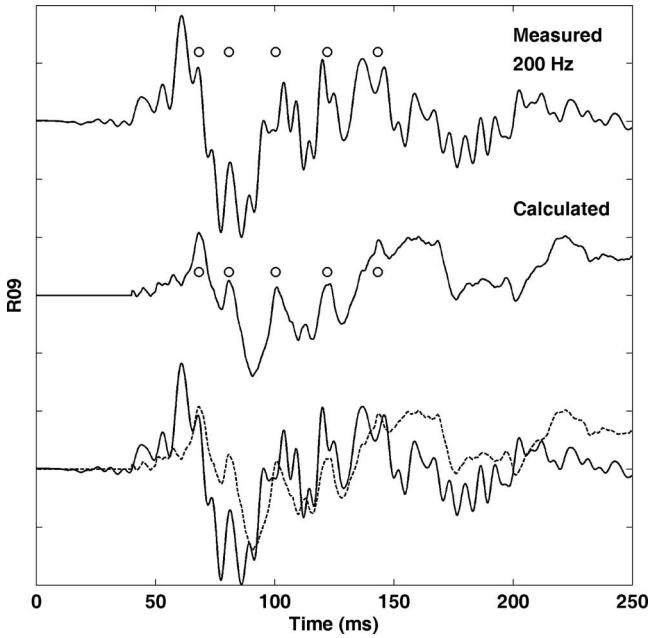


FIG. 11. Comparison of measured and theoretical waveforms for a source at SP1 and receiver at location 9, behind building A. The measured waveform (top) has been low-pass filtered at 200 Hz. The calculated waveform (center) has been muted for times before 40 ms, and time-shifted by 2.0 ms to align with the measured waveform. Open circles indicate arrival times identified from the finite difference wavefield movie as horizontal diffractions around two corners of building A, followed by arrivals along the same path after single or multiple reflections in the street between buildings A and E. The calculated (dashed line) and measured (solid line) waveforms are plotted together at the bottom of the figure.

sence of diffractions for the path of the particular arrivals that arrive at a sensor. The largest time shifts were applied to NLOS sensors. The propagation path for the first arrival at sensor 16 was diffracted around the corner of building E, very early in its travel to the sensor, so its propagation velocity was lower than estimated from Fig. 5 (for propagation without diffraction), soon after it left the source, and thus a larger time difference is expected for this particular sensor. With these adjustments, Fig. 10 shows that the overall waveform shapes and arrival times generally agree at many of the sensor locations.

The two-dimensional FDTD calculations omit acoustic propagation paths over the rooftops. Liu and Albert (2006) found that this two-dimensional calculation method produced good agreement with waveform measurements conducted using a single-storey right-angle wall as the scattering obstacle. In Fig. 9(a), the worst agreement between the calculated and measured peak amplitudes was for sensor 9 located behind building A. The measured and calculated waveforms for this position are compared in Fig. 11. Here, the measured waveform has been low-pass filtered at 200 Hz. The calculated waveform was muted (zeroed) for times before 40 ms and time-shifted by 2.0 ms so that the peak amplitude aligns with the *second* measured peak. Careful study of the wavefield movie produced by the finite difference calculations allowed the calculated wave arrivals to be identified. The small amplitude arrivals at 45 ms (and following) were produced by waves propagating through and reverberating within the building itself. The first large arrival (at 68 ms) was produced by the wave that diffracted around the two building corners, followed by large arrivals at 80, 101, 122, and 144 ms that followed the same path after single or multiple reverberations in the street between buildings A and E. These theoretical arrival times are shown as open circles in Fig. 11, and they seem to agree well with some pulses on the measured waveform, given the uncertainty in the acoustic wave speed, as discussed earlier.

The horizontal ray diffracts around corners of 66° and 90°, and arrives at the sensor after traveling 24.1 m. For the propagation path over the roof, the ray grazes the first roof edge, then is diffracted through 66° and 57° to reach the sensor, with a path length of 21.0 m. The 3 m path length difference corresponds to an arrival time difference of about 8 ms (compared to 7 ms between the first and second positive pulses on the measured waveform), with the rooftop arrival expected to arrive first. A rough calculation of the insertion loss produced by the building for the 100 Hz octave band (Kang, 2006, page 151) gives 18 and 16 dB for the horizontal and rooftop paths, respectively, close to the amplitude difference shown in Fig. 9. In this case, the rooftop path is the largest expected arrival, but it is not included in the theoretical calculations, and this omission explains the differences in the amplitude and waveform comparisons.

The waveform agreement shown in Fig. 10 is acceptable, considering all of the approximations in the computational model. These include the two-dimensional formulation that neglects rooftop wave propagation paths, the approximate corrections for the variable wave velocity in air because of nonlinear effects, and the simplified buildings used in the model (ignoring the façade details, etc.).

A similar waveform comparison is shown for SP4 in Fig. 12. In this case, the propagation distances were all about the same, ranging from 86 to 117 m, so a constant wave velocity for air of 375 m s^{-1} was used in these calculations. The measured waveforms have been low-pass filtered, and a time shift ranging from 9 to 23 ms was applied to some of the calculated waveforms. No time shift was applied to sensors 6, 15, 16, and 17. In some cases, the time shift applied to the calculated traces were selected to align the minimum (negative) pressure rather than the peak positive pressure. Many, but not all of the measured waveform details, appear on the calculated time series for SP4.

Some of the calculated waveforms for SP4 show good agreement with the measurements, including sensors 3, 8, 18, and 20 (LOS). The initial propagation paths to these sensors are all relatively "open," without path obstructions near the sensor. Most of the other sensors have calculated initial arrivals smaller than the measured amplitudes (sensors 2, 4, 6, 9, 12, and 15). The calculated waveforms for sensors 9 and 15, located on the back wall of buildings, again show the poorest agreement with the measured waveforms. Sensor 17 on building G also shows poor agreement with the measured waveform. The high pressure blast sensor 1, selected to measure the C4 blast waves at a distance of 1 m (SP 1), was unable to detect the much lower waveform produced by the explosive charge SP4. For some of the sensors (for example, sensor 12), the waveform agreement is better later in the time series, and the worst agreement is for the first arrival.

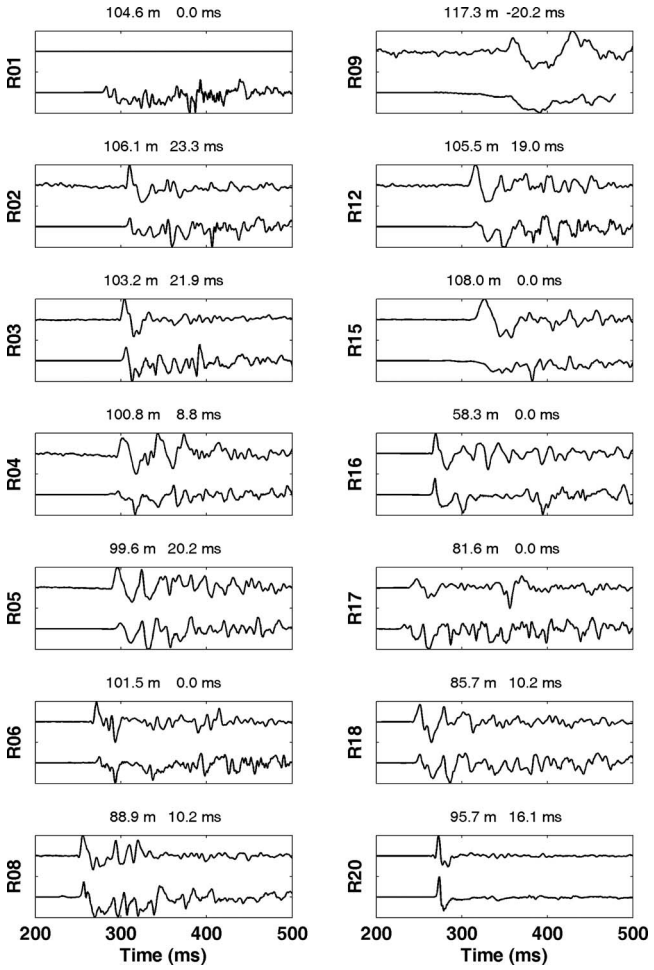


FIG. 12. Measured (upper) and theoretical (lower) waveforms at 14 locations in the artificial village for source point 4. The observed waveforms have been low-pass filtered at 200 Hz, and the calculated waveforms have been time-shifted to correct for variable acoustic velocity as explained in the text. The side labels give the locations, as shown in Fig. 1, and these waveforms are all individually normalized.

IV. CONCLUDING REMARKS

Measurements of acoustic pulse propagation in a complex urban environment demonstrate that, for propagation over straight, line-of-sight propagation paths, complex high frequency waveforms comprised of multiple reflections are recorded, with higher amplitudes than for open air situations. In non-line-of-sight situations, while the rigid ground and building surfaces tend to preserve the high frequency waveforms and produce little attenuation, diffraction introduces low-pass filtered arrivals, producing very complex waveforms. In all cases, the waveforms display more complexity and longer time durations than for open air situations.

Finite difference calculations using a simplified two-dimensional representation of the urban propagation environment were conducted on a personal computer and produced good agreement with the measured peak amplitudes. The calculation method also gave good agreement with many waveform characteristics. The major discrepancies between the calculations and measurements occurred for sensors located directly behind buildings.

The added complexity of the acoustic waveforms produced in an area with many buildings will have a large effect on the performance of acoustic sensor systems. It can be difficult for human listeners to recognize that playbacks of some of the complex signals shown in Fig. 3 were produced by acoustic pulses; instead, they often sound more like jet noise.

Methods using direction-of-arrival or beam-forming estimates to determine an acoustic source location (Haykin, 1985) are likely to perform poorly, especially in NLOS situations, because the sound will arrive at the sensor from many different directions, with possibly none of those directions correctly pointing toward the actual source location (sensor location 18 and source point 1 in Fig. 1 are examples of this situation). Even in LOS situations (for example, sensor location 5 in Fig. 1), bearing estimates may vary widely as multiple reflections with different angles of arrival reach the sensor, and only a general quadrant direction is likely to be produced. However, if the geometry of the situation is known, time-reversal processing methods can potentially be used to decode the complex signatures and produce a correct source location (Albert *et al.*, 2005).

ACKNOWLEDGMENTS

We thank the late Mr. Rich Andrejkovics and Mr. Andre Edwin, SYDET Project Officer of PM-CCS, for funding the experimental measurements presented here. This paper is dedicated to Rich's memory. The analysis was funded by the U.S. Army Corps of Engineers Engineering Research and Development Center research program. We also thank the many people who assisted with the field measurements, including David Carbee, Stephen Decato, and Dr. Joyce Nagle (ERDC-CRREL); Chris Reiff and Herb Brann (Army Research Laboratory); Gary Leadore, George Eason, Thomas Ruffing, Patrick Whaling, and James Aguiar (Aberdeen Test Center); Ron Akers, Stan Ellis, George Eason, and others from Fort Benning. In addition, Major Joseph Haydon (U.S. Army) assisted with the measurement design, Dr. Joyce Nagle conducted a review of the urban acoustic literature used in this paper, and David Leese (ERDC-ITL) provided meteorological data. We also thank two anonymous reviewers for useful comments that improved the manuscript.

Albert, D. G., Liu, L., and Moran, M. (**2005**). "Time reversal processing for source location in an urban environment," J. Acoust. Soc. Am. **118**, 616–619.

Alberts, W. C. K., Noble, J. M., and Coleman, M. A. (**2008**). "Sound propagation in the vicinity of an isolated building: An experimental investigation," J. Acoust. Soc. Am. **124**, 733–742.

Wyle Laboratories (**1971**). *Community Noise*, prepared for the U.S. Environmental Protection Agency, Washington, DC.

ANSI (**1983**). "Estimating airblast characteristics for single point explosions in air, with a guide to evaluation of atmospheric propagation and effects," American National Standards Institute, S2-20, pp. 1–25.

Berenger, J. P. (**1994**). "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200.

Blumrich, R., and Heimann, D. (**2002**). "A linearized Eulerian sound propagation model for studies of complex meteorological effects," J. Acoust. Soc. Am. **112**, 446–455.

Botteldooren, D. (**1994**). "Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid," J. Acoust. Soc. Am. **95**, 2313–2319.

Brown, A. L., and Lam, K. C. (**1987**). "Levels of ambient noise in Hong Kong," Appl. Acoust. **20**, 85–100.

Bullen, R., and Fricke, F. (**1976**). "Sound propagation in a street," J. Sound Vib. **46**, 33–42.
Garcia, A., and Faus, L. J. (**1991**). "Statistical analysis of noise levels in urban areas," Appl. Acoust. **34**, 227–247.
Hahn, C. D., and Leese, D. L. (**2002**). "Automated environmental data collection at Fort Benning, Georgia, from May 1999 to July 2001," Engineer Research and Development Center, Vicksburg, MS.
Haykin, S., ed. (**1985**). *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Heutschi, K. (**1995**). "A simple method to evaluate the increase of traffic noise emission level due to buildings, for a long straight street," Appl. Acoust. **44**, 259–274.
Iu, K. K., and Li, K. M. (**2002**). "The propagation of sound in narrow street canyons," J. Acoust. Soc. Am. **112**, 537–550.
Kang, J. (**2000**). "Sound propagation in street canyons: Comparison between diffusively and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404.
Kang, J. (**2001**). "Sound propagation in interconnected urban streets: A parametric study," Environ. Plan. B: Plan. Des. **28**, 281–294.
Kang, J. (**2006**). *Urban Sound Environment* (Taylor & Francis, London).
Lee, K. P., and Davies, H. G. (**1975**). "Nomogram for estimating noise propagation in urban areas," J. Acoust. Soc. Am. **57**, 1477–1480.
Liu, L., and Albert, D. G. (**2006**). "Acoustic pulse propagation near a right-angle wall," J. Acoust. Soc. Am. **119**, 2073–2083.
Oldham, D. J., and Radwan, M. M. (**1994**). "Sound propagation in city streets," Build. Acoust. **1**, 65–88.
Picaut, J., Polles, T. L., L'Hermite, P., and Gary, V. (**2005**). "Experimental study of sound propagation in a street," Appl. Acoust. **66**, 149–173.
Picaut, J., Simon, L., and Hardy, J. (**1999**). "Sound field modeling in streets with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645.
Pierce, A. D. (**1989**). *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).
Schroder, C. T., and Scott, W. R. (**2002**). "On the stability of the FDTD algorithm for elastic media at a material interface," IEEE Trans. Geosci. Remote Sens. **40**, 474–481.
Van Renterghem, T., Salomons, E., and Botteldoorn, D. (**2006**). "Parameter study of sound propagation between city canyons with a coupled FDTD-PE model," Appl. Acoust. **67**, 487–510.
Virieux, J. (**1984**). "SH-wave propagation in heterogeneous media: Velocity-stress finite difference method," Geophysics **49**, 1933–1942.
Wang, S. (**1996**). "Finite-difference time-domain approach to underwater acoustic scattering problems," J. Acoust. Soc. Am. **99**, 1924–1931.
Wiener, F. M., Malme, C. I., and Gogos, C. M. (**1965**). "Sound propagation in urban areas," J. Acoust. Soc. Am. **37**, 738–747.
Yee, K. S. (**1966**). "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307.

Solving the vibroacoustic equations of plates by minimization of error on a sample of observation points

Olivier Collery[a)]
AIRBUS Operations SAS, Department of Acoustics and Environment, 316 route de Bayonne, F-31060 Toulouse Cedex 09, France

Jean-Louis Guyader
Laboratoire Vibrations Acoustique, INSA Lyon, Bâtiment St Exupéry, 25 bis avenue Jean Capelle, F-69621 Villeurbanne Cedex, France

(Received 12 February 2009; revised 16 October 2009; accepted 13 December 2009)

In the context of better understanding and predicting sound transmission through heterogeneous fluid-loaded aircraft structures, this paper presents a method of solving the vibroacoustic problem of plates. The present work considers fluid-structure coupling and is applied to simply supported rectangular plates excited mechanically. The proposed method is based on the minimization of the error of verification of the plate vibroacoustic equation of motion on a sample of points. From sampling comes an aliasing effect; this phenomenon is described and solved using a wavelet-based filter. The proposed approach is validated in presenting very accurate results of sound radiation immersed in heavy and light fluids. The fluid-structure interaction appears to be very well described avoiding time-consuming classical calculations of the modal radiation impedances. The focus is also put on different samplings to observe the aliasing effect. As perspectives sound radiation from a non-homogeneous plate is solved and compared with reference results proving all the power of this method. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290982]

PACS number(s): 43.40.Rj, 43.40.Dx [DF] Pages: 1347–1356

I. INTRODUCTION

An aircraft structure is strongly heterogeneous: It can be considered as an assembly of stiffened curved panels with non-uniform skin thickness. It includes beams to support the floor as well as large size windows in the flight deck. The turbulent boundary layer excitation is one of the main sources of aircraft interior noise over a large frequency range. In the industrial context of better understanding and predicting sound transmission through heterogeneous fluid-loaded aircraft structures, a formulation is proposed with purpose of assessing quickly sound radiation from these complex aircraft structures. In particular, the present paper aims at presenting a method applied to radiation from simply supported baffled plates exposed to a half-space of fluid and excited mechanically.

Sound radiation from plates has been widely studied for many years since this simple problem can explain how and why structures radiate noise. The calculation of radiated pressure leads to complexity and is time consuming even if the structure is a plate, e.g., the aircraft applications proposed by Graham[1,2] and Maury *et al.*[3,4] The need of computationally simpler method is obvious. Nevertheless, the drastic simplification of neglecting fluid loading is not considered because it is not appropriate in lot of practical situations.

In this paper, modal decomposition of plate vibrations into *in vacuo* modes, also called dry modes, is considered. This approach is widely used in literature. Fluid-structure interaction is characterized through the modal radiation impedances, which are the time-consuming terms of the calculation. A large number of papers have put its focus on their evaluations.[5–11] Conclusions drawn from these studies have emphasized difficulties to calculate these impedances and their necessity in heavy fluid. Some models predicting sound radiation from plates with fluid loading have been proposed in literature.[12–17] A state of the art on the prediction of radiation from plates is presented by Atalla and Nicolas.[18] Mutual-radiation resistances are often considered as having negligible impact on radiated sound power, but more practically this simplification leads to a strong reduction in computation time since each mode remains uncoupled. More recently, Li and Gibeling[19] and Sha *et al.*[20] presented simple calculation methods for the mutual radiation resistances. In particular, Li and Gibeling[19] showed that they may not be meaningless as compared with the self-terms in a quite wide frequency range.

A choice has to be made between approximate models, easy to use and quick to compute but giving only tendencies, or accurate models limited by the calculation times and consequently not adapted to large parametric studies, hence the constant need of new tools for predicting accurately and rapidly vibrations and sound radiation from plates. The formulation presented by Atalla and Nicolas[18] has this objective. Their approach is based on a Taylor's expansion of Green's function, which brings some real improvement.

The present work is led within this context. It still uses displacement expansion into series of *in vacuo* plate eigenfunctions but instead of improving modal radiation imped-

[a)]Author to whom correspondence should be addressed. Electronic mail: olivier.collery@airbus.com

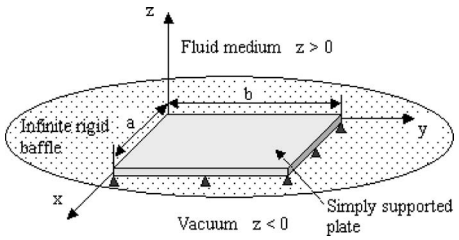


FIG. 1. Geometry of the problem.

ances it solves directly the plate equation of motion coupled to the acoustic media.

The formulation proposed is based on sampling and modal optimization. Indeed, modal amplitudes are optimized so as to satisfy the governing equation for a sampling of observation points of the plate under fluid loading. For the sake of brevity the method is called MESOP (minimization of error on a sample of observation points) in the following. This approach has been adopted due to the existence of numerous advanced algorithms treating efficiently of optimization problems.

It is shown in this paper that MESOP offers an efficient tool for assessment of sound radiation from simply supported baffled plates immersed in light or heavy fluid and a special focus is also put on presenting possibilities of extension given by the method. In Sec. II, homogeneous plates are considered and corresponding theoretical analysis associated is derived from plate sampling and modal decomposition. Then aliasing phenomenon and its solving through a wavelet-based filter are described. Finally the resolution principle is presented.

Section III is devoted to numerical validations. The model is first validated in heavy fluid. The fluid-structure coupling appears to be very well described. Different plate samplings are considered to observe the aliasing effect. Next a numerical validation is made in light fluid and a comparison in term of computation time is led on different algorithms. The present approach shows here all its interest.

Finally the model is generalized in Sec. IV to nonhomogeneous plates. The particular case of a plate with varying thickness is highlighted and used to give credit to the generalized formulation.

II. MESOP PRINCIPLE AND THEORY ASSOCIATED

A. Statement of the problem

The structure of interest is a simply supported thin rectangular plate inserted in a baffle (Fig. 1). The baffle is assumed to be infinite, plane, and rigid. The plate-baffle system separates vacuum ($z<0$) from an acoustic medium ($z>0$). As a result, the plate radiates into the fluid in the semi-infinite space. The plate produces flexural vibrations in response to a harmonic mechanical force per unit area of angular frequency ω and applied normally to the plate surface. Transverse shear and rotary inertia effects are neglected. A harmonic motion is considered but for sake of simplicity time dependence $e^{j\omega t}$ is applied. Transverse motion $W(x,y)$ of the homogeneous plate is derived using Love–Kirchhoff assumptions as follows:

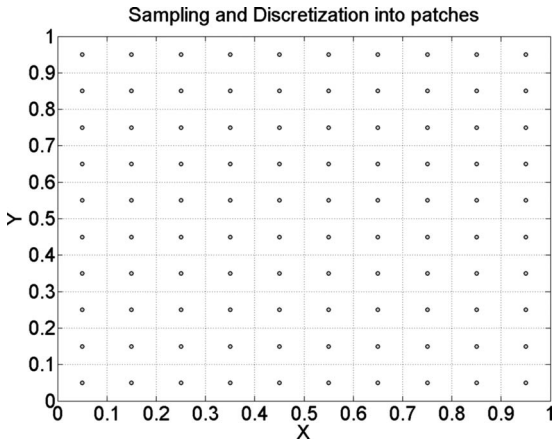


FIG. 2. Two linked approximations: discretization into patches (——) and observation on samples (○).

$$D\Delta^2 W(x,y) - \omega^2 \rho h W(x,y) = F(x,y) - p_0(x,y), \tag{1}$$

where D is the flexural rigidity, ρ is the mass density, h is the thickness, $F(x,y)$ is the driving force per unit area, and $p_0(x,y)$ is the fluid loading pressure. $p_0(x,y)$ is defined by the well-known Rayleigh integral

$$p_0(x,y) = \int_{S_p} j\rho_0 \omega V(x',y') \frac{e^{-jk_0 R}}{2\pi R} dx'dy', \tag{2}$$

where S_p is the surface of the plate, $k_0=\omega/c_0$ is the acoustic wave number, ρ_0 and c_0 are the density and the sound speed in the fluid, and $R^2=(x-x')^2+(y-y')^2$.

The vibroacoustic response of the plate is assessed in terms of three vibroacoustic indicators, namely, the mean square velocity, the radiated sound power, and the radiation factor (see definitions in Appendix B).

B. Sampling approach

To model the vibroacoustic behavior of the plate, a sampling approach is used. This method consists in observing the verification of the equation of motion on a sample of points of the plate. Equation (1) is valid for any point (x,y). Therefore, combining Eq. (1) and the sampling approach results in the following equation:

$$D\Delta^2 W(x_i,y_i) - \omega^2 \rho h W(x_i,y_i) = F(x_i,y_i) - p_0(x_i,y_i), \quad \forall\, i \in [1, N_x \times N_y], \tag{3}$$

where (x_i,y_i) are the coordinates of sample point i, and N_x and N_y are the numbers of samples in the x and y directions, respectively.

The limitation to a sample of points verifying the equation of motion will naturally produce only an approximation of the solution. To go further it is necessary to express the boundary pressure with the plate displacement.

C. Fluid loading

The fluid loading pressure $p_0(x,y)$ is calculated using the collocation method (see Appendix A). Thus, a discretization of the plate into rectangular patches is made (Fig. 2). These patches are chosen so as to let patch centers corre-

spond to sample points. Thus, same approximation level is used in the plate modeling and the fluid loading calculation.

The pressure on the plate is as a result defined by

$$p_0(x_i,y_i) = \sum_{j=1}^{N_x\times N_y} Z_{ij}V(x_j,y_j) = j\omega \sum_{j=1}^{N_x\times N_y} Z_{ij}W(x_j,y_j). \quad (4)$$

The impedance terms Z_{ij} are given by

$$Z_{ij} = \frac{p_0(x_i,y_i)}{V(x_j,y_j)} = \frac{1}{2\pi}\rho_0 j\omega \frac{e^{-jk_0 d_{ij}}}{d_{ij}} S_j, \quad (5)$$

$$Z_{ii} = \frac{p_0(x_i,y_i)}{V(x_i,y_i)} = \rho_0 c(1-e^{-jk_0 r_i}), \quad (6)$$

where r_i is the radius of the circular patch of surface S_i.

The equation of the plate coupled with the acoustic medium is thus given by

$$D\Delta^2 W(x_i,y_i) - \omega^2\rho h W(x_i,y_i) + j\omega \sum_{j=1}^{N_x\times N_y} Z_{ij}W(x_j,y_j) = F(x_i,y_i), \quad \forall\, i \in [1,N_x\times N_y]. \quad (7)$$

As seen before, a second approximation appears here due to the finite size of pistons used in the collocation technique. It is implicitly related to the sample points located at the pistons midpoints.

D. Modal decomposition

To solve the problem, thanks to modal optimization, modal decomposition of plate displacement is introduced. Let the flexural motion of the plate be expanded into series of *in vacuo* modes

$$W(x,y) = \sum_{m,n} a_{mn}\Phi_{mn}(x,y), \quad (8)$$

where a_{mn} and Φ_{mn} are the modal amplitude and the modal shape, respectively, of the (m,n) mode. For the case of a simply supported plate, Φ_{mn} is given by

$$\Phi_{mn}(x,y) = \sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}, \quad (9)$$

where a and b represent the plate dimensions in the x and y directions. Note that expansion (8) does not necessitate that functions Φ_{mn} are modal shapes but only basis functions. Nevertheless the physical nature of mode shapes helps the convergence and when mode shapes are known, like in this case, their use is very convenient compared to other basis. In addition, it can be pointed out the perspective to treat other boundary conditions in just switching the functional basis, e.g., clamped plates. An approach coupled to a finite element method (FEM) can also be envisaged for complex boundary conditions representative of aircraft applications.

Using modal decomposition in the equation of motion given by Eq. (7) leads to a set of equation written at each sample point that depends on the modal amplitudes

$$D\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Delta^2\Phi_{mn}(x_i,y_i) - \omega^2\rho h\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Phi_{mn}(x_i,y_i) + j\omega\sum_{j=1}^{N_x\times N_y} Z_{ij}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Phi_{mn}(x_j,y_j) = F(x_i,y_i), \quad \forall\, i \in [1,N_x\times N_y], \quad (10)$$

where $m_{\max}$ and $n_{\max}$ are the maximal orders of the modes in the x and y directions, respectively. For the sight of simplicity, the equation of motion can be rewritten as

$$\sum_{q=1}^{q_{\max}}\left[D\Delta^2\Phi_q(x_i,y_i) - \omega^2\rho h\Phi_q(x_i,y_i) + j\omega\sum_{j=1}^{N_x\times N_y} Z_{ij}\Phi_q(x_j,y_j)\right]a_q = F(x_i,y_i), \quad \forall\, i \in [1,N_x\times N_y], \quad (11)$$

$$[\mathbf{S}]\{\mathbf{A}\} = \{\mathbf{F}\}, \quad (12)$$

where q is the mode index, $[\mathbf{S}]$ is the sampling matrix of size equal to the number of samples times the number of modes, $\{\mathbf{A}\}$ is the modal amplitude vector of size equal to the number of modes, and $\{\mathbf{F}\}$ is the driving force per unit area vector of size equal to the number of samples. After plate sampling the modal truncation represents the second type of approximation of the method described here. Therefore, the maximal orders of the modes will be considered as a parameter of investigation.

E. The aliasing phenomenon

In the present method, the plate displacement is expressed from mode responses. Let us consider that $M\times M$ modes are used. In the wave number domain each plate mode is defined by a couple of wave numbers of the form $(m\pi/a, n\pi/b)$ where $m=1, M$ and $n=1, M$. Thus, all wave numbers used to describe the plate response are located in a rectangular domain (Fig. 3).

Let us consider a same number N of sample points in both directions. The sample of observation points is regularly distributed over the plate at points of coordinates $(pa/N, qb/N)$ where a/N and b/N are the spacings between samples in x and y directions and $N\times N$ is the number of sample points considered. In the wave number domain this sampling corresponds to a couple of wave numbers of the form $(p\pi/a, q\pi/b)$ where $p=1,N$ and $q=1,N$. Plate displacements observed in the method contain all wave numbers located in a rectangular domain (Fig. 3).

Three configurations are possible depending on the number of modes $M\times M$ and the number of sample points $N\times N$.

- $M>N$. In this case (Fig. 3) modes' wave number components remain uncontrolled by observation points and the method failed to predict vibroacoustic plate response. This

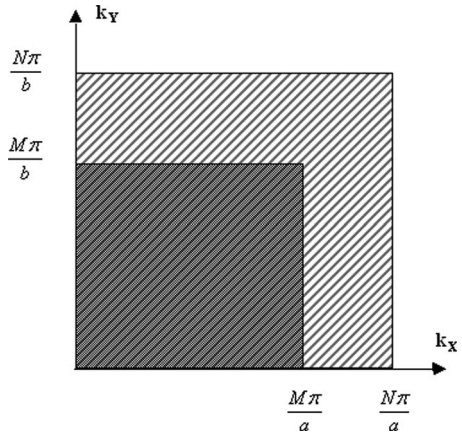


FIG. 3. Wave number domains of the modes $M \times M$ and sample points $N \times N$. Case of $M < N$.

case corresponds to an underdetermined problem: less equations (sample points) than unknowns (modal amplitudes).

- $M = N$. All modes' wave numbers are here exactly controlled by sample points and the vibroacoustic response is well described.
- $M < N$. In this case the aliasing problem is observed, the problem is also called overdetermined. As explained in signal processing textbooks, the signal observed by sampling has wave number components not taken into account by modal decomposition. The aliasing effect can be negligible is some cases (see Sec. III B 2) but leads to wrong modal amplitudes in most cases and as a result to wrong vibroacoustic response.

F. Solving the aliasing problem by using a wavelet-based filter

To avoid aliasing, all wave number contributions that cannot be taken into account by modal decomposition must be suppressed. This is achieved by filtering high wave number components not controlled by modal decomposition. In this paper, a two-dimensional (2D) filtering at level 1 based on Daubechies wavelets of order 10 is applied to Eq. (12) before resolution.

More precisely, vector $\{\mathbf{F}\}$ and each column of matrix $[\mathbf{S}]$ are filtered in order to retain only low wave number components corresponding to wave numbers of modes used in the calculation. Concerning vector $\{\mathbf{F}\}$ the filtering distributes more equally the force on the plate, the force level is therefore lowered at its primary location as illustrated in Fig. 4 where f_C is the critical frequency of the plate considered.

Matrix $[\mathbf{S}]$ is made of N rows and M columns, where N is the number of observation points and M is the number of modes considered. This matrix has to be filtered consequently for each column, that is, for the spatial distribution of observation points of the plate at one given order (m,n) of the functional basis. In that purpose, each column is first reshaped in a 2D-matrix and the matrices obtained are then filtered.

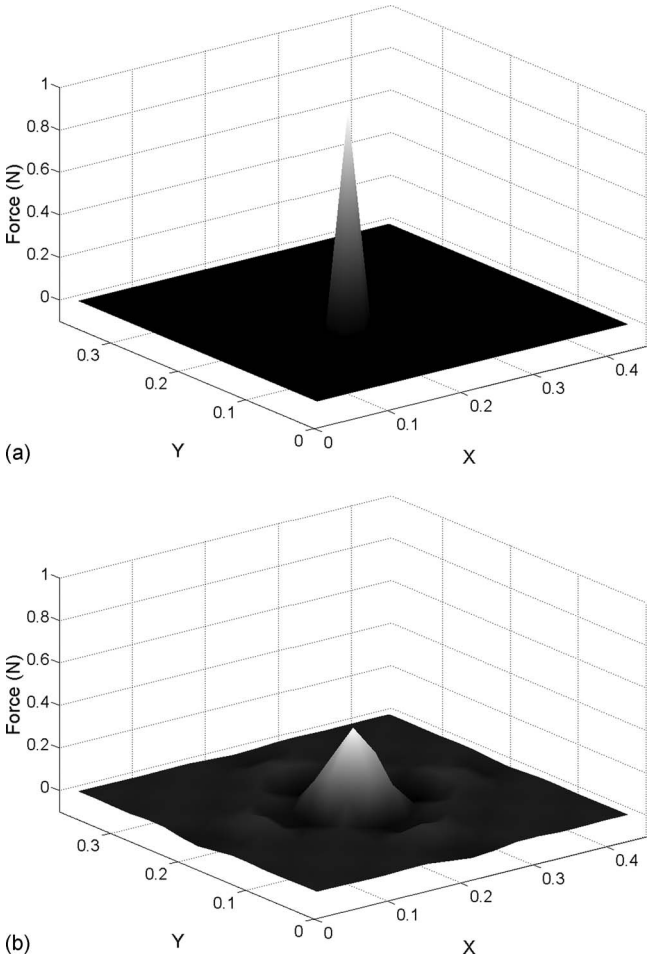


FIG. 4. Filter impact on force distribution. Case of sound radiation into water from a simply supported plate ($a/b=1.2$, $a/h=455$, $f_C>20$ kHz) driven by a point force at its center: (a) without filtering and (b) with filtering.

G. Resolution

To solve the problem the modal amplitudes must satisfy Eq. (12). However, in general, it is impossible and one has to find optimum modal amplitudes satisfying Eq. (12) as much as possible.

Several possibilities to find optimum modal amplitude values exist, generally through the minimization of residual function. Many algorithms performing this computation are available in the market. The present work will use the basic inverse function for square sampling matrix $[\mathbf{S}]$, i.e., number of modes equal to number of sample points. Nevertheless, in most cases, the sampling matrix is rectangular and Eq. (12) is solved in the least squares sense using LAPACK routines implemented in MATLAB; one gets the following minimization:

$$\min_{\{\mathbf{A}\}} \|[\mathbf{S}]\{\mathbf{A}\} - \{\mathbf{F}\}\|_2. \tag{13}$$

This algorithm will be compared in terms of computation time to other algorithms implemented in MATLAB such as the pseudoinverse function, the derivative-free method, or genetic algorithms.

Minimization yields optimum modal amplitude values, which are finally used to assess the three vibroacoustic indi-

cators. Note that a larger number of sample points than modes corresponds to most practical cases, then the filtering allows decreasing the size of matrix system (12) and therefore the associated resolution process of Eq. (13).

III. NUMERICAL VALIDATION FOR HOMOGENEOUS PLATES IN HEAVY AND LIGHT FLUIDS

To validate the proposed theory, the following vibroacoustic indicator is assessed: the radiated sound power W. Its definition is given in Appendix B. The radiated sound power is expressed in decibel with a reference level equal to 1. The validation of the present model is divided into two parts: sound radiation first into heavy fluid and then into light fluid.

Two input data need to be defined before computation: the number of samples and the number of modes considered. They are strongly responsible of accurate results. Convergence criteria will be consequently defined too. To illustrate aliasing effect and for the sake of simplicity, the same number of modes (M) is used in both directions. The same assumption is made with the number of samples (N).

A. Vibroacoustic response of plates under heavy fluid loading

The heavy fluid yields here a strong fluid-structure interaction. As a consequence, the method is really tested in a difficult case.

Experimental results under heavy fluid loading are not common in literature. Such experiments are difficult to set up; in fact, only few reference results exist but none was found allowing comparison with. However, some numerical results of sound radiation of simply supported baffled plates into water exist.[14,15,17] Berry and Foin *et al.* dealt with this problem using a variational formulation solved by the Rayleigh–Ritz method, whereas Laulagnet solved analytically this problem in calculating directly the radiation impedances. Note that the state-space approximation[21,22] permits to determine the fluid-loaded resonances through classical analyses (e.g., eigenvalue decomposition).

A numerical validation is led in heavy fluid comparing the optimized results with Berry's work in the case of a simply supported plate $455\times375\times1$ mm^3 immersed in water ($\rho_0=1000$ kg m^{-3} and $c_0=1500$ m s^{-1}) and driven by a point force F_p at its center [$F_p=\|F_p\|\delta(x-a/2)\delta(y-b/2)$ and $\|F_p\|=1$]. Using sampling approach, one gets the following expression of the driving force per unit area:

$$F(x_i,y_i)=\begin{cases}\dfrac{\|F_p\|}{S_i} & \text{if sample } i \text{ is sample excited}\\ 0 & \text{else.}\end{cases} \tag{14}$$

The frequency range of study is 10–4000 Hz. A large number of *in vacuo* plate modes are considered to be sure that no modal contribution is missing.

To investigate the influence of sampling, comparison is made in the three following configurations: $M=N$, $M>N$, and $M<N$.

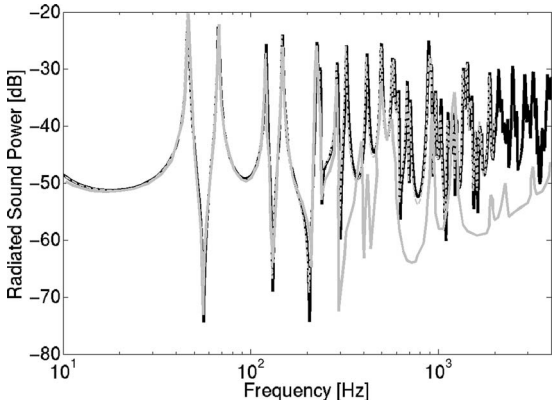


FIG. 5. Radiation into water by the proposed method with $M=N$ (black solid line) compared to Berry's results (gray solid line) and FEM results (gray dashed line). Case of a simply supported plate ($a/b=1.2$, $a/h=455$, $f_C>20$ kHz) driven by a point force at its center.

1. Calculation using as many modes as samples ($M=N$)

The resolution is quite simple since the sampling matrix is square in this case and therefore a basic inverse function leads to the solution. Figure 5 presents comparative results with those from Berry. Comparison shows a good accuracy until 300 Hz between optimized results and results from Berry. Beyond 300 Hz, a different behavior is observed. A cross validation with a commercial FEM software is consequently led proving the validity of our results. The FEM model size limits results to 2 kHz. Note that the MESOP computation runs around 70 times faster than the FEM one. As a consequence a poor convergence of Berry's results is assumed; indeed only 169 functions were used in its Rayleigh–Ritz expansion of the plate displacement whereas the present method uses 1369 functions ($M=37$). Note also that Berry uses a strong truncation to compute radiation impedance coefficients. This reference case treats only of subsonic modes because of the high value of the sound speed in water ($c_0=1500$ m s^{-1}) yielding a very high critical frequency ($f_C>20$ kHz). A case handling supersonic modes will be investigated in Sec. IV.

2. Calculation using more modes than samples ($M>N$)

With $M>N$, the sampling matrix becomes rectangular and Eq. (12) is solved in the least squares sense in optimizing Eq. (13). Results are compared to those of the previous case $M=N$ (Fig. 6). A poor agreement is observed proving the incapacity of the method to solve underdetermined problem, i.e., more unknowns (modal amplitudes) than samples.

3. Calculation using less modes than samples ($M<N$)

In this case, the sampling matrix is also rectangular and Eq. (12) is solved in the least squares sense in optimizing Eq. (13). Results are also compared to those of the case $M=N$ (Fig. 6). Moreover this case must lead to the aliasing phenomenon. The wavelet-based filter is consequently needed.

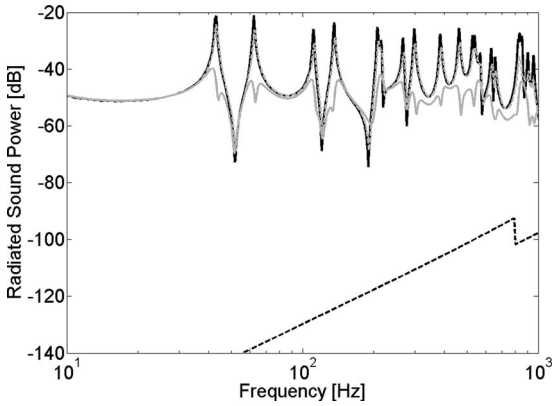


FIG. 6. Radiation into water by the proposed method with $M=N$ (black solid line), $M>N$ (black dashed line), $M<N$ without filtering (solid gray line), and with filtering (dashed gray line). Case of a simply supported plate ($a/b=1.2$, $a/h=455$, $f_C>20$ kHz) driven by a point force at its center.

To observe the influence of this filter, two calculations have been done: without and with filtering both sampling matrix $[\mathbf{S}]$ and vector $\{\mathbf{F}\}$.

On the one hand results without filtering show clearly problems of convergence due to aliasing. The global behavior is relatively well described while resonances are strongly underestimated.

On the other hand results with filtering show all the power of filtering. In fact, resonances are well described and aliasing effect is totally suppressed.

Finally, the validation into heavy fluid has proved the very good convergence of the method in a difficult case. The fluid-structure interaction is very well described. In addition to this, the mutual influence of both parameters has been illustrated: modes considered and plate sampling. The aliasing effect has been underlined when the sampling matrix is overdetermined ($M<N$) and the filter chosen has demonstrated its capacity to suppress this undesirable effect.

4. Evaluation of the radiation impedances and aliasing filter impacts on computation time

The impact on resolution time of these two MESOP steps is illustrated in Table I through the Berry's case described above. Two configurations are treated: $M=N=37$ and $M=37$ with $N=15$. Results show first of all the interest to consider the configuration $M<N$ even if it presents an aliasing effect. In fact, this effect is easily treated with the filter proposed and allows a faster computation with a gain of 85% here on the resolution time. The evaluation of the radiation impedances is naturally decorrelated of the configuration since it only depends on the number of samples considered.

TABLE I. Computation time of different MESOP steps for one frequency point depending on the choice of M and N. Simulation time obtained on a 700-Mflops computer

	$M=N=37$	$M=37/N=15$
Radiation impedances	0.63s	0.63s
Aliasing filter	0.00s	0.30s
Resolution	1.41s	0.19s
Total	2.04s	1.12s

The evaluation time of these impedances becomes therefore non-negligible when the more efficient configuration $M<N$ is treated; e.g., the radiation impedances represent here 30% of the total computation time when $M=N$ and 56% when $M<N$.

5. Criteria defining the number of modes and samples needed

The choice of the modes considered is made in applying the following criterion (15) based on *in vacuo* modes. One gets all modes placed in a quadrant depending on the maximal frequency of study and a constant γ. The constant γ is taken to be equal to 2 in heavy fluid in order to take into account added mass effect compared to vacuum. In the present paper, modes are of the form $M\times M$, and therefore $M=\max(m,n)$ must be considered.

$$\left\{(m,n)\ \text{such as}\ \left(\frac{m}{a}\right)^2+\left(\frac{n}{b}\right)^2\leq\gamma\sqrt{\frac{4\rho h}{\pi^2 D}}f_{\max}\right\}. \quad (15)$$

Concerning the influence of sampling, it is obvious that the more sample points taken, the greater accuracy. A sampling criterion of $\lambda_{\mathrm{fl}}/4$ is enough to assess accurately the vibroacoustic response below the critical frequency f_C, where the flexural wavelength λ_{fl} is given by

$$\lambda_{\mathrm{fl}}=\frac{\sqrt{\omega}}{f}\sqrt[4]{\frac{D}{\rho h}}. \quad (16)$$

Beyond the critical frequency, the acoustic wavelength is more restrictive and has to be considered.

B. Vibroacoustic response of plates under light fluid loading

In this case several results are available in literature for comparison. The method is applied to a steel plate $1000\times1000\times1$ mm^3, driven by a point force at (0.7, 0.2) of 1 N. The following properties of air are chosen to simulate the light fluid loading: $\rho_0=1.2$ kg m^{-3} and $c_0=340$ m s^{-1}. The structural loss factor is taken to be $\eta=0.01$. This factor is introduced in the complex Young modulus E^* such as

$$E^*=E(1+j\eta). \quad (17)$$

The three vibroacoustic indicators of interest are plotted between 10 Hz and 1 kHz (note that the corresponding critical frequency is 12 kHz). Simulation results are compared to reference results from C-Valor-test cases collection.[23]

1. Validation

To validate the present approach into light fluid, 20×20 modes and 41×41 samples are used ($M<N$) and the filter is applied. Results presented in Fig. 7 show an excellent agreement between optimized and reference results over the complete frequency range of interest. The present method allows an accurate assessment of the response and the radiation of the plate studied. The number of modes to be consid-

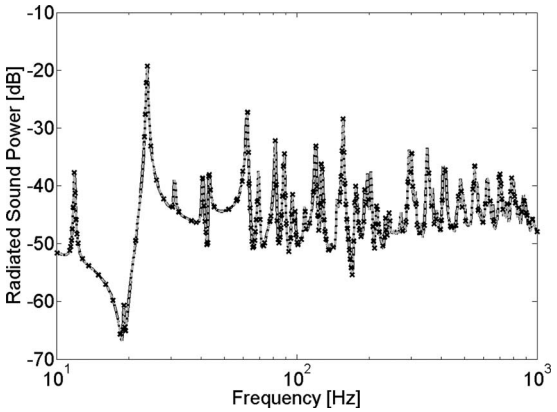


FIG. 7. Radiation into air by the proposed method with $M<N$ without filtering ($\times$) and with filtering (dashed line) compared to reference results from C-Valor (solid line). Case of a simply supported thin plate ($a/b=1$, $a/h=1000$, $f_C=12$ kHz) driven by a point force at $(7a/10, b/5)$.

ered in light fluid is defined using a constant γ equal to 1 in Eq. (15) when the added mass effect is negligible compared to vacuum.

2. Aliasing effect into light fluid

The aliasing effect is then investigated into light fluid. In that purpose, the vibroacoustic response of the previous plate is calculated using same configuration ($M<N$) without filtering the sampling matrix nor the driving force per unit area vector. Results are compared to the case $M=N(=41)$ presenting no aliasing (Fig. 7). It can be shown that aliasing effect is negligible into light fluid. Indeed a very good agreement is observed between both results. The reason is that *in vacuo* modes taken to model the plate behavior are very lowly modified by the light fluid loading and, as a consequence, aliasing effect becomes negligible in that particular case. In fact, Φ_{mn} can be considered as modal shapes for this case and not only basis functions.

C. Choice of the optimization algorithm

Several algorithms are available to solve the problem defined by Eq. (13) with $M<N$. This problem corresponds to the case where the sampling matrix $[\mathbf{S}]$ is rectangular, i.e., the most common case. The following algorithms implemented in MATLAB have been tested: least squares algorithm using LAPACK routines, the pseudoinverse function, the derivative-free method, and the genetic algorithms. To help us choose the best optimization algorithm, a comparison in terms of accuracy and computation time has been performed.

A preliminary computation has permitted to eliminate the derivative-free method and the genetic algorithms; both are not adapted because the function to optimize needs too much computation time. Indeed both methods are based on iterations and each iteration corresponds to an evaluation of the function to optimize.

Comparison between the least squares algorithm and the pseudoinverse technique has been led on the previous plate in light fluid between 10 and 100 Hz with 450 frequency points. 12×12 sample points and 6×6 modes are considered. The computation of vibroacoustic indicators using least squares algorithm represents the reference in terms of accuracy since it was used to validate the method. Results are illustrated in Fig. 8. Comparison has demonstrated the better efficiency of the first algorithm. In fact, both algorithms have an excellent agreement but the second one takes around four times longer than the first one to compute (Table II).

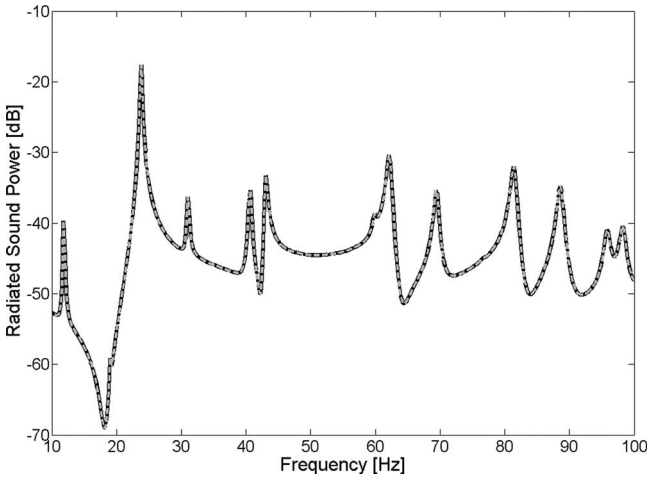


FIG. 8. Comparison in terms of radiated sound power between the least squares algorithm (dashed line) and the pseudoinverse function (solid line). Case of a simply supported thin plate ($a/b=1$, $a/h=1000$, $f_C=12$ kHz) driven by a point force at $(7a/10, b/5)$.

IV. EXTENSION TO NON-HOMOGENEOUS PLATES IN LIGHT FLUID

A. Generalization

This section aims at solving vibroacoustic problems using MESOP method for non-homogeneous plates. The present approach can be easily generalized to this case, since just the governing equation has to be changed. Non-homogeneous plate equation for bending motion according to Love–Kirchhoff assumptions is given by

$$D\Delta^2 W - \omega^2 \rho h W + 2\frac{\partial D}{\partial x}\frac{\partial(\nabla^2 W)}{\partial x} + 2\frac{\partial D}{\partial y}\frac{\partial(\nabla^2 W)}{\partial y} + (\nabla^2 D) \times(\nabla^2 W) - (1-\nu)\left(\frac{\partial^2 D}{\partial y^2}\frac{\partial^2 W}{\partial x^2} - 2\frac{\partial^2 D}{\partial y\,\partial x}\frac{\partial^2 W}{\partial y\,\partial x} + \frac{\partial^2 D}{\partial x^2}\frac{\partial^2 W}{\partial y^2}\right) = F - p_0, \tag{18}$$

where ν is the Poisson ratio. Compared to the homogeneous plate case, the difference comes from additional stiffness terms. Note that this equation is only valid for a spatial variation of the Young modulus E, the thickness h, and the Poisson ratio ν, which are very common in aircraft structures.

TABLE II. Comparison of two optimization algorithms. Computation time of 450 frequency points over 10–100 Hz. Simulation time obtained on a 700-Mflops computer.

	Least squares algorithm	Pseudoinverse technique
Computation time (s)	6.7	26.1

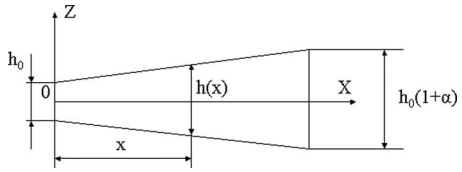


FIG. 9. Plate with linearly varying thickness.

Next the MESOP approach is applied in a straightforward manner, Eq. (7) of the homogeneous plate case being modified by introducing the new plate equation with additional terms. An example is proposed below.

B. Example of varying thickness

To validate MESOP approach for non-homogeneous plate, prior work of Miloudi *et al.*[24] is used for comparison. Miloudi *et al.* put their focus on the influence of varying thickness on sound radiation from a simply supported baffled plate into air. The studied thickness is only varying in the x direction (Fig. 9) as follows:

$$h(x) = h_0\left(1 + \frac{\alpha}{a}x\right), \tag{19}$$

where h_0 is the thickness at $x=0$ and α is a given constant. As a consequence, Eq. (18) becomes

$$D\Delta^2 W(x,y) - \omega^2 \rho h(x) W(x,y) + 2\frac{\partial D(x,y)}{\partial x}\frac{\partial(\nabla^2 W(x,y))}{\partial x} + \frac{\partial D^2(x,y)}{\partial x^2}\left(\frac{\partial W^2(x,y)}{\partial x^2} + \nu\frac{\partial W^2(x,y)}{\partial y^2}\right) = F(x,y) - p_0(x,y). \tag{20}$$

The varying thickness is assumed to have no influence on the radiation impedances estimated by the collocation method. Using the MESOP approach leads to Eqs. (21) and (22) to be solved where $[\mathbf{S_{NH}}]$ is the sampling matrix for this non-homogeneous case. Note that $\{\Phi_{mn}(x,y)\}$ still form a functional basis but are of course not modal shapes of the heterogeneous plate.

$$\begin{aligned} &D\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Delta^2\Phi_{mn}(x_i,y_i) \\ &\quad - \omega^2\rho h(x_i)\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Phi_{mn}(x_i,y_i) \\ &\quad + 2\frac{\partial D}{\partial x}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\frac{\partial(\nabla^2\Phi_{mn}(x_i,y_i))}{\partial x} \\ &\quad + \frac{\partial D^2}{\partial x^2}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\left(\frac{\partial^2\Phi_{mn}(x_i,y_i)}{\partial x^2} + \nu\frac{\partial^2\Phi_{mn}(x_i,y_i)}{\partial y^2}\right) \\ &\quad + j\omega\sum_{j=1}^{N_x\times N_y} Z_{ij}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} a_{mn}\Phi_{mn}(x_j,y_j) \\ &= F(x_i,y_i), \quad \forall\, i \in [1, N_x \times N_y], \end{aligned} \tag{21}$$

$$[\mathbf{S_{NH}}]\{\mathbf{A}\} = \{\mathbf{F}\}. \tag{22}$$

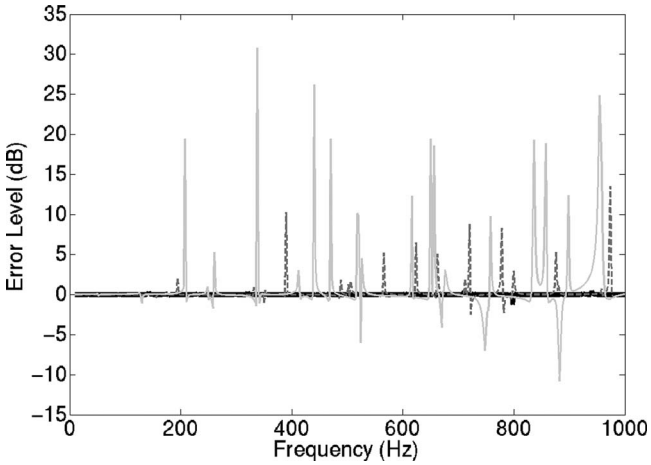


FIG. 10. Error level of the radiated sound power when $M<N$ with and without filtering. Case of a simply supported thick plate ($a/b=1$, $a/h=125$, $f_C=1.6$ kHz) with varying thickness driven by a point force at (0.33,0.38): $\alpha=0$ (black solid line), $\alpha=0.2$ (gray dashed line), and $\alpha=1$ (gray solid line).

The radiation factor (see Appendix B) of three steel plates 1×1 m^2 with $h_0=8$ mm and $\alpha=0$, 0.2, and 1, all driven by a point force at (0.33,0.38) of 1 N, is here investigated. The following results are calculated using $N=35$ and $M=12$ after filtering system (22). It can be observed that the greater alpha factor, the stronger aliasing influence, as illustrates Fig. 10 plotting the error level of mean square velocity with and without filtering. In fact, *in vacuo* modes used for homogeneous plates into light fluid become less and less valid when the plate becomes heterogeneous.

The present method allows an excellent agreement with results from Miloudi *et al.*'s work over 10 Hz–3 kHz, as illustrated in Fig. 11 where the case of higher structural heterogeneity, that is, $\alpha=1$, is treated. This observation gives credit to the generalized model, which seems to be also powerful with non-homogeneous plates.

In addition, this case shows the very good accuracy of the method proposed and the aliasing filter for supersonic modes, that is, for frequencies beyond the critical frequency; i.e., $f>1.6$ kHz here. All previous cases were validated for

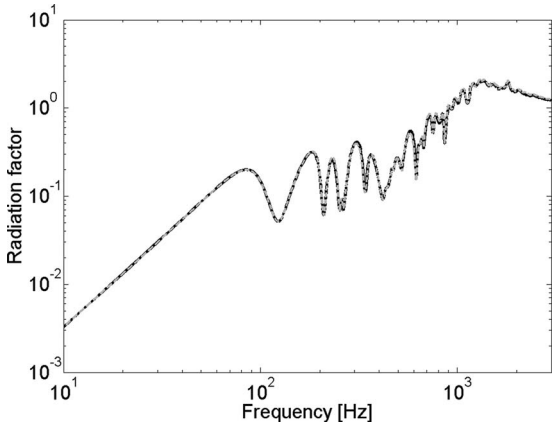


FIG. 11. Radiation factor into air by the proposed method (dashed line) compared to results from Miloudi *et al.* (solid line). Case of a simply supported thick plate ($a/b=1$, $a/h=125$, $f_C=1.6$ kHz) with varying thickness ($\alpha=1$) driven by a point force at (0.33,0.38).

cases handling only subsonic modes. The reasons were the following: On the one hand the critical frequency is too high in water because of the sound speed ($c_0=1500$ m s^{-1}) around 4.5 times higher than in air; i.e., $f_C^{\text{water}} \approx 20 \times f_C^{\text{air}}$. On the other hand the homogeneous cases immersed in air do not permit to conclude since the aliasing effect is then negligible.

Finally, one can also notice here a real advantage of this approach compared to classical ones, that is, the simplicity of changing of plate structures in just modifying the differential equation of motion.

V. CONCLUSION

An innovative formulation based on sampling and optimization has been presented. Its simplicity through the use of powerful existing algorithms was attractive, and comparison with reference results and literature in heavy and light fluids has demonstrated the power of this approach in terms of simplicity and computation time. After a plate sampling into observation points, modal optimization permits to predict accurately and above all very quickly sound radiation from simply supported baffled plates excited mechanically. The fluid-structure coupling is precisely included without neglecting mutual radiation resistances whereas the difficulties, both theoretical and numerical, of calculating the modal radiation impedances are avoided. Hence all the interest of this method.

Furthermore, the generalization to non-homogeneous plates investigated here allows one to think about new applications as sandwich plates or more generally heterogeneous structures. Extension to different boundary conditions has been emphasized and different excitations can also be envisaged. In fact, the sampling approach gives the possibility to implement deterministic excitations such as the plane wave.

Finally, the MESOP approach represents not only a powerful tool to get accurate and quick results of sound radiation from fluid-loaded structures but also a method adapted to aircraft specificities.

APPENDIX A: COLLOCATION METHOD

The collocation method used to define the radiation impedances necessary to calculate the fluid loading pressure is here described. The Rayleigh integral formulated in Eq. (2) is first split up as two integrals depending on the observation point. And then, using the discretization of the surface S_p into patches, one gets at Q_0 (point inside the patch i)

$$p_0(Q_0) = \sum_{j=1, j\neq i}^{N} \int_{S_j} j\omega\rho_0 V(Q_j) \frac{e^{-jk_0\|Q_jQ_0\|}}{2\pi\|Q_jQ_0\|} dQ_j + \oint_{S_i} j\omega\rho_0 V(Q_i) \frac{e^{-jk_0\|Q_iQ_0\|}}{2\pi\|Q_iQ_0\|} dQ_i, \tag{A1}$$

where Q_i and Q_j are points of patches i and j, respectively. The patches are assumed to be small enough so that the behavior of any point inside a patch can be approximated by its center value, that is to say the corresponding sample point. The velocity V is thus considered to be uniform inside each patch and the distance between two points from different patches is assumed to be equal to the distance between the two centers. The first integral of Eq. (A1) is consequently rewritten as

$$\int_{S_j} j\omega\rho_0 V(Q_j) \frac{e^{-jk_0\|Q_jQ_0\|}}{2\pi\|Q_jQ_0\|} dQ_j = \frac{1}{2\pi} j\omega\rho_0 V(x_j,y_j) \frac{e^{-jk_0 d_{ij}}}{d_{ij}} S_j, \tag{A2}$$

where d_{ij} is the distance between patches i and j, and S_i is the surface of the patch i.

Next, the singularity of the Rayleigh integral in case of patch's radiation on itself (i.e. emitting patch=receiving patch) can be avoided using polar coordinates in order to integrate the singularity. In that purpose, a circular patch of the same surface as the rectangular one of interest is considered; it yields

$$\oint_{S_i} j\omega\rho_0 V(Q_i) \frac{e^{-jk_0\|Q_iQ_0\|}}{2\pi\|Q_iQ_0\|} dQ_i = \rho_0 c V(x_i,y_i)(1 - e^{-jk_0 r_i}), \tag{A3}$$

where r_i is the radius of the circular patch of surface S_i.

From these results, the radiation impedance Z_{ij} corresponding to the ratio of the fluid loading pressure at patch i and the velocity at patch j is introduced and can be derived as follows:

$$Z_{ij} = \frac{p_0(x_i,y_i)}{V(x_j,y_j)} = \frac{1}{2\pi} j\omega\rho_0 \frac{e^{-jk_0 d_{ij}}}{d_{ij}} S_j, \tag{A4}$$

$$Z_{ii} = \frac{\rho_0(x_i,y_i)}{V(x_i,y_i)} = \rho_0 c(1 - e^{-jk_0 r_i}). \tag{A5}$$

APPENDIX B: VIBROACOUSTIC INDICATORS

The radiated sound power W is defined by

$$W = \frac{1}{2} \int_x \int_y \text{Re}(p_0(x,y) V^*(x,y)) dxdy, \tag{B1}$$

where Re and $*$ denote the real part and the conjugation, respectively. Using the sampling approach, which assumes a uniform behavior around each observation point, one gets a discrete formulation of the radiated sound power

$$W = \frac{1}{2} \sum_{i=1}^{N_x \times N_y} \text{Re}(p_0(x_i,y_i) V^*(x_i,y_i)) S_i. \tag{B2}$$

In the same way, the following definition of the mean square velocity $\langle V^2 \rangle$ can be formulated using plate sampling:

$$\langle V^2 \rangle = \frac{1}{2S_p} \int_x \int_y |V(x,y)|^2 dxdy \tag{B3}$$

and

$$\langle V^2 \rangle = \frac{1}{2S_p} \sum_{i=1}^{N_x \times N_y} |V(x_i,y_i)|^2 S_i. \tag{B4}$$

Finally, the radiation factor σ is defined by

$$\sigma = \frac{W}{\rho_0 c \langle V^2 \rangle S_p}. \tag{B5}$$

[1]W. R. Graham, "High frequency vibration and acoustic radiation of fluid-loaded plates," Philos. Trans. R. Soc. London, Ser. A **352**, 1–43 (1995).
[2]W. R. Graham, "Boundary layer induced noise in aircraft, part i: The flat plate model," J. Sound Vib. **192**, 101–120 (1996).
[3]C. Maury, P. Gardonio, and S. J. Elliott, "A wavenumber approach to modelling the response of a randomly excited panel, part i: General theory," J. Sound Vib. **252**, 83–113 (2002).
[4]C. Maury, P. Gardonio, and S. J. Elliott, "A wavenumber approach to modelling the response of a randomly excited panel, part ii: Application to aircraft panels excited by a turbulent boundary layer," J. Sound Vib. **252**, 115–139 (2002).
[5]G. Maidanik, "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826 (1962).
[6]A. J. Price and M. J. Crocker, "Sound transmission through double panels using statistical energy analysis," J. Acoust. Soc. Am. **47**, 683–693 (1970).
[7]I. L. Ver and C. I. Holmer, *Noise and Vibration Control*, edited by L. L. Beranek (McGraw-Hill, New York, 1971), pp. 287–296.
[8]C. E. Wallace, "Radiation resistance of rectangular panel," J. Acoust. Soc. Am. **51**, 946–952 (1972).
[9]L. D. Pope and R. C. Leibowitz, "Intermodal coupling coefficients for a fluid-loaded rectangular plate," J. Acoust. Soc. Am. **56**, 408–415 (1974).
[10]B. E. Sandman, "Fluid-loaded vibration of an elastic plate carrying a concentrated mass," J. Acoust. Soc. Am. **61**, 1503–1510 (1977).
[11]J. P. Coyette, G. Lielens, M. Robbé, and P. Neple, "An efficient method for evaluation diffuse field joint acceptance functions for cylindrical and truncated conical geometries," J. Acoust. Soc. Am. **117**, 1009–1019 (2005).
[12]H. G. Davies, "Low frequency random excitation of water-loaded rectangular plates," J. Sound Vib. **15**, 107–126 (1971).
[13]R. F. Keltie and H. Peng, "The effects of modal coupling on the acoustic power radiation from panels," Trans. ASME **109**, 48–54 (1987).
[14]A. Berry, "A new formulation for the vibrations and sound radiation of fluid-loaded plate with elastic boundary conditions," J. Acoust. Soc. Am. **96**, 889–901 (1994).
[15]B. Laulagnet, "Sound radiation by a simply supported unbaffled plate," J. Acoust. Soc. Am. **103**, 2451–2462 (1998).
[16]H. Nelisse, O. Beslin, and J. Nicolas, "A generalized approach for the acoustic radiation from a baffled or unbaffled plate with arbitrary boundary conditions, immersed in a light or heavy fluid," J. Sound Vib. **211**, 207–225 (1998).
[17]O. Foin, J. Nicolas, and N. Atalla, "An efficient tool for predicting the structural acoustic and vibration response of sandwich plates in light or heavy fluid," Appl. Acoust. **57**, 213–242 (1999).
[18]N. Atalla and J. Nicolas, "A new tool for predicting rapidly and rigorously the radiation efficiency of plate-like structures," J. Acoust. Soc. Am. **95**, 3369–3376 (1994).
[19]W. L. Li and H. J. Gibeling, "Determination of the mutual radiation resistances of a rectangular plate and their impact on the radiated sound," J. Sound Vib. **229**, 1213–1233 (2000).
[20]K. Sha, J. Yang, and W.-S. Gan, "A simple calculation method for the self- and mutual-radiation impedance of flexible rectangular patches in a rigid infinite baffle," J. Sound Vib. **282**, 179–195 (2005).
[21]J. A. Giordano and G. H. Koopman, "State space boundary element-finite element coupling for fluid-structure interaction analysis," J. Acoust. Soc. Am. **98**, 363–372 (1995).
[22]K. A. Cunefare and S. D. Rosa, "An improved state-space method for coupled fluid-structure interaction analysis," J. Acoust. Soc. Am. **105**, 206–210 (1999).
[23]C-Valor, "Benchmark for vibroacoustic software validation," SFM-SFA (French societies for mechanics and acoustics), 1996.
[24]A. Miloudi, N. Hamzaoui, and J.-L. Guyader, "Influence of thickness defects on the acoustical radiation of plates," Acta Acust. Acust. **85**, 213–220 (1999).

Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds

Jin Yong Jeon, Pyoung Jik Lee,[a] and Jin You
Department of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

Jian Kang
School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

(Received 21 March 2009; revised 30 September 2009; accepted 4 January 2010)

In this study, urban soundscapes containing combined noise sources were evaluated through field surveys and laboratory experiments. The effect of water sounds on masking urban noises was then examined in order to enhance the soundscape perception. Field surveys in 16 urban spaces were conducted through soundwalking to evaluate the annoyance of combined noise sources. Synthesis curves were derived for the relationships between noise levels and the percentage of highly annoyed (%HA) and the percentage of annoyed (%A) for the combined noise sources. Qualitative analysis was also made using semantic scales for evaluating the quality of the soundscape, and it was shown that the perception of acoustic comfort and loudness was strongly related to the annoyance. A laboratory auditory experiment was then conducted in order to quantify the total annoyance caused by road traffic noise and four types of construction noise. It was shown that the annoyance ratings were related to the types of construction noise in combination with road traffic noise and the level of the road traffic noise. Finally, water sounds were determined to be the best sounds to use for enhancing the urban soundscape. The level of the water sounds should be similar to or not less than 3 dB below the level of the urban noises.
 [DOI: 10.1121/1.3298437]

PACS number(s): 43.50.Rq, 43.50.Sr, 43.50.Qp [BSF] Pages: 1357–1366

I. INTRODUCTION

Exposure-effect relationships have been used to describe an environmental situation and to predict the effect of changes from exposure. The relationships between the exposure to noise from a single source type, such as aircrafts, road traffic, or railways, and the annoyance based on a large international data set have been reported.[1,2] Most of the previous studies applied the percentage of a population expressing high annoyance such as %HA (percentage of highly annoyed) and %A (percentage of annoyed) since Schultz proposed the concept of %HA.[1–3] The %HA is the percentage of annoyance responses that exceed a defined cutoff point. Schultz[4] used a cutoff of 72 on a scale from 0 to 100 in his synthesis to define %HA. The %A was also defined as responses in the upper 50% of the annoyance scale. Recently, some researchers have investigated the total annoyance that is caused by combined noise sources because community noises are seldom heard and are concurrent.[5–12] In addition, methods for quantifying the total annoyance caused by the exposure of a dwelling noise from combined sources were proposed. However, most of these studies only investigated transportation sources and did not deal with other urban noises such as construction noise.

Most studies evaluated the annoyance in the dwellings of the residents. The ISO 15666 standard describes the proper wording of the questions for the field survey as "at home," which means within the residents' home or outside the home, for example, in the garden or on the balcony.[13] However, outdoor environments such as public parks and plazas are also important. Contrary to home environments, the evaluation of outdoor environments is rather complicated because of the interactions between various sound sources and other acoustical and nonacoustical factors. Therefore, the evaluation of the outdoor environment has been conducted in this study based on the concept of soundscape.

The initial concept of soundscape was proposed as an attempt to construct an analytical perspective that would describe the total acoustic environment over time and across cultures.[14] Therefore, most soundscape studies concern the qualitative analysis of soundscapes. Schulte-Fortkamp and Fiebig[15] adopted the grounded theory as a sociological approach, and Berglund and Nilsson[16] proposed a tool for measuring soundscape quality by attribute profiling. Raimbault[17] also conducted qualitative judgments of an urban soundscape with nine semantic scales. However, the methods for evaluating a soundscape are different according to the purposes of the studies and the researchers, and thus it is difficult to directly compare the results of these studies. Even though many recent studies adopted the soundwalking methodology to identify the perception of an urban acoustic environment,[16,18] the standardization of the procedures for assessing soundscapes are still being discussed in the ISO TC43 SC1 WG 54 (perceptual assessment of soundscape quality).

The initial approach of soundscape studies focused on ecologically improving soundscapes as a method to reduce

[a]Author to whom correspondence should be addressed. Electronic mail: pyoungjik@daum.net

pollution. More recently, physical characteristics and mental perceptions of the auditory environment have been studied using soundscape approaches.[19] As Schulte-Fortkamp and Fiebig[15] as well as Dubois *et al.*[20] suggested, it is important to establish the interactive relationships between the sound environment and the human recognition of the positive aspects of a soundscape. In the engineering field, many studies on soundscape have attempted to control urban noises in terms of enhancing the sound environment. These attempts include the methods for physical measurement, subjective evaluation, and architectural design strategies to improve the soundscape in urban public spaces.[21] One of the proposed solutions to improve the soundscape in urban spaces is to introduce natural sounds to mask urban noises such as road traffic and construction noises. Although classifying the characteristics of a soundscape in an open public space is important, the active treatment of the sound environment is often needed in order to improve the acoustical quality. Among the various kinds of natural sounds, the sound of water from water features such as fountains was found to be the most applicable in parks or city squares to provide pleasantness.[21] This is one of the reasons why fountains or sculptures with water streams are easily found in open public spaces of many cities. Watts *et al.*[22] studied characteristics of water drop sounds in terms of the noise masking effects. However, little is known about the effective sound level of water sounds in various situations of urban soundscapes.

In this study, the noise annoyance of typical urban spaces that had construction as well as road traffic noises was investigated using field surveys through soundwalking. The methods used in previous community noise studies were adopted, such as the %HA and %A. Analysis was also made using the semantic scales for evaluating the soundscape quality. Laboratory auditory experiments were then conducted in order to quantify the total annoyance that is caused by both construction and road traffic noises. The methodologies were then applied to the actual conditions of urban soundscape, which included the sounds of water along with other noise sources. Sound preference tests were conducted with varying levels of water sounds, considering individual and combined noise sources. The results of the experiment provided optimum combinations of water sounds and urban noises.

II. FIELD SURVEY THROUGH SOUNDWALKING

A. Methodology

1. Site selection

Field surveys were performed in 16 urban sites in Seoul and Bundang (the biggest satellite city of Seoul). These sites were chosen based on the fact that the dominant noise sources in the area were both road traffic noise and construction noise, and because the number of complaints concerning the noise from the construction work had significantly increased in recent years. Eight sites were located in residential areas, and the other sites were in open public spaces such as the streets, parks, and squares. The selected construction sites had different noise sources (pile driving, drill rig, excavator, hammering, and so on) according to the construction procedure.

2. Evaluation of soundscapes

In this study, the evaluation of the soundscapes is based on both a quantitative approach and a qualitative approach through soundwalking. The exposure-effect relationship is derived from the quantitative approach using a general method of analysis that is linked with community noise. The use of semantic scales will provide various subjective perceptions including comfort and pleasantness as well as annoyance, and can also be applied to the process of characterization of urban soundscapes. The exposure-effect relationship can be explained by the results of the qualitative analysis, and the subjective perceptions can also be connected with the annoyance level.

Soundwalking was employed as a method to evaluate and describe the urban sound environment. In previous studies,[16,18] soundwalking was carried out along a number of sites; however, the soundwalking in this study was conducted at each site separately. It was conducted in silence, and the participants were asked to concentrate on what they could hear as they walked and observed the urban environment. After soundwalking for 30 min at each site, the participants were asked to evaluate the annoyance of the noise sources and the quality of the soundscape.

All the field surveys were conducted in the afternoon (13:00–18:00) based on the assumption that outdoor activities are most frequent during this period. The field survey continued for 4 days (four sites per day), with 15 subjects (seven female and eight male) aged 20–30. All the subjects participated all the soundwalks at the same time; thus their exposure would be about the same.

3. Questionnaire

The ICBEN team 6 recommends the use of two questions to measure annoyance reactions in order to be able to make comparisons between social surveys, so that both a 5-point verbal scale and an 11-point numerical scale were used for the quantitative analysis of the urban soundscapes in terms of annoyance.[13] The questions addressed in the ISO 15666 were translated into Korean and the standardized noise annoyance modifiers (in Korean) were also used in the five-point verbal scale questions.[23]

The questionnaire included questions to assess road traffic and construction noises, as well as general questions about the participants themselves. The questions were arranged in three basic sections. The first section sought to obtain their annoyance of the noise sources, which contained three questions that aimed to assess their overall impression of the sound environment and to obtain responses to the road traffic and construction noises separately. In the second section, ten semantic scales describing the quality of the soundscape were used for qualitative judgments of the urban soundscapes. The ten semantic scales were comfort, quiet, harmonious, soft, monotonous, weak, light, pleasant, unique, and warm. They were selected based on a review of previous studies on semantic differential descriptors of sound environments.[16,17] A pair of descriptive adjectives defined the semantic words used, with ten words on the right and their antonyms on the left. On a seven-point scale, three points

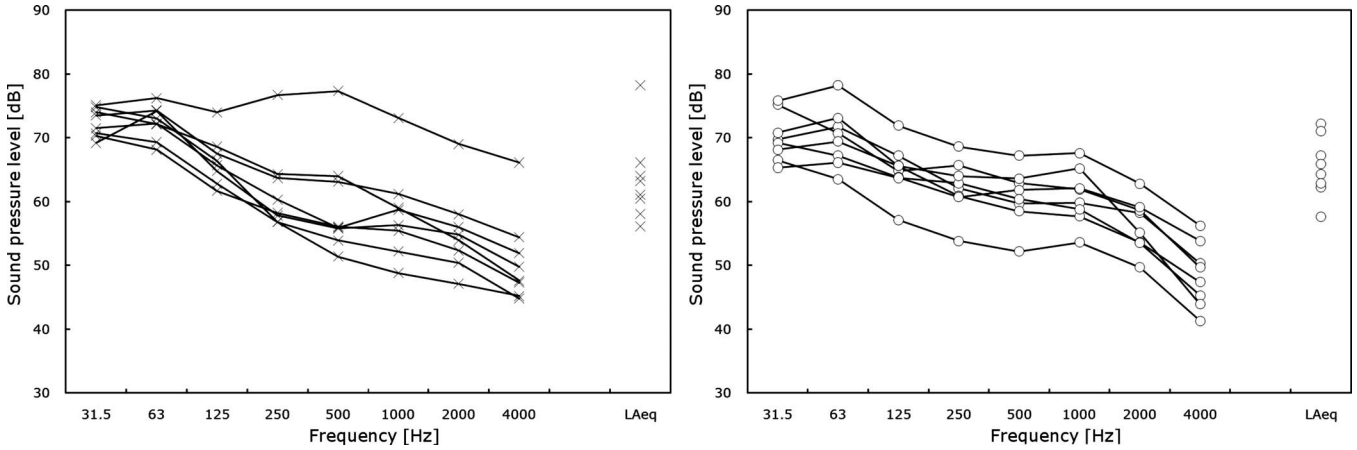


FIG. 1. Spectra of 30-min measurements conducted in eight residential areas (left) and in eight open public spaces (right).

mean the participant had a very strong impression of the soundscape. The third section dealt with demographic data (age and gender), dwelling type, health condition, acoustic comfort, and loudness. Acoustic comfort and loudness to the urban soundscape were evaluated using an 11-point numerical scale.

4. Noise measurement

In this study, a 30-min L_{Aeq} was measured by one subject during the soundwalk. Thus, the recording corresponds with the actual exposure. The sound pressure level (SPL) measurement was conducted using a binaural microphone (B&K type 4101), and the video recording was also made using a camcorder (Sony DCR-HC90) in order to present visual information of the soundscape during the auditory test (see Secs. III and IV).

The measured spectra are shown in Fig. 1. Low frequency components were dominant in most sites, but one site had significant high frequency components due to the hammering on metal and motor vehicles and this occurred in an open public space. The 30-min L_{Aeq} of the case study sites ranged from 54 to 78 dB.

B. Results and discussions

1. Quantitative analysis results

Annoyance responses to road traffic and construction noises were obtained from a 5-point verbal scale and an 11-point numerical scale that were later translated into a scale from 0 to 100 for the assessment of %HA and %A. The exposure-response relationships that were obtained as a function of L_{Aeq} from the 5-point verbal and 11-point numerical scales are shown in Fig. 2. The curves presented in Fig. 2 were obtained through quadratic ordinary least-squares regression. Questions in the 11-point numerical scale caused less annoyance than the 5-point verbal scale in terms of %HA. It is interesting to note that the participants rarely chose "8," "9," and "10" in the 11-point scale even though they were exposed to higher noise levels than noise level of "7" point they selected. The curve from the 5-point verbal scale shows that the %HA is 0 at around 60 dB but the curve from the 11-point numerical scale shows that the %HA is 0 at around 63 dB.

The %A curves from the 5-point scale and the 11-point scale were almost the same, which is different from the %HA. This indicates that the percentage of responses that select "3," "4," and "5" in the 5-point scale is almost same to

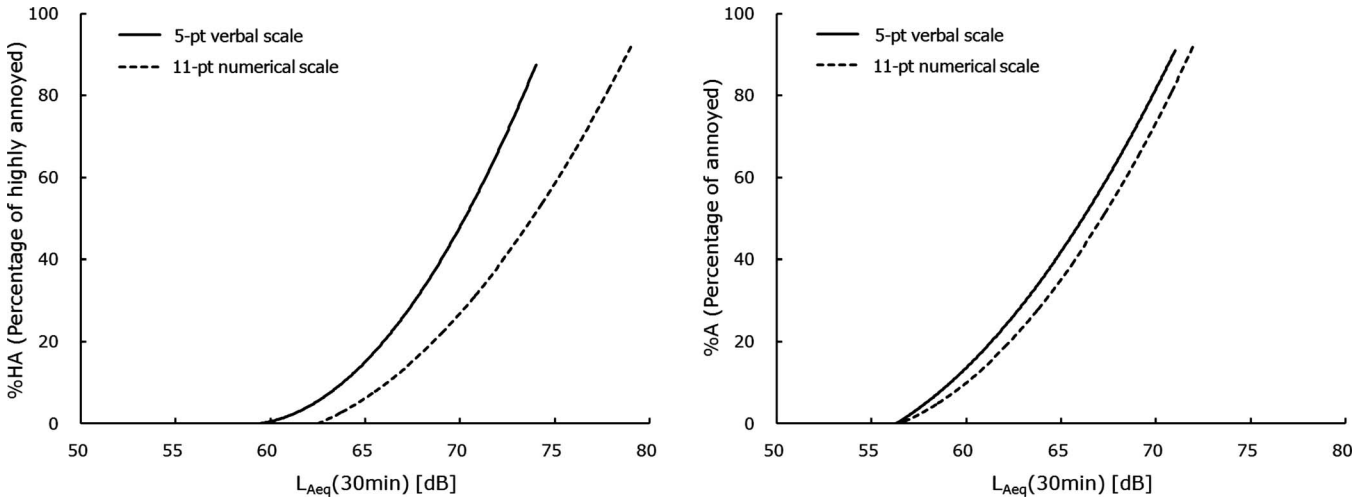


FIG. 2. The percentages highly annoyed (%HA) and annoyed (%A) as a function of L_{Aeq}.

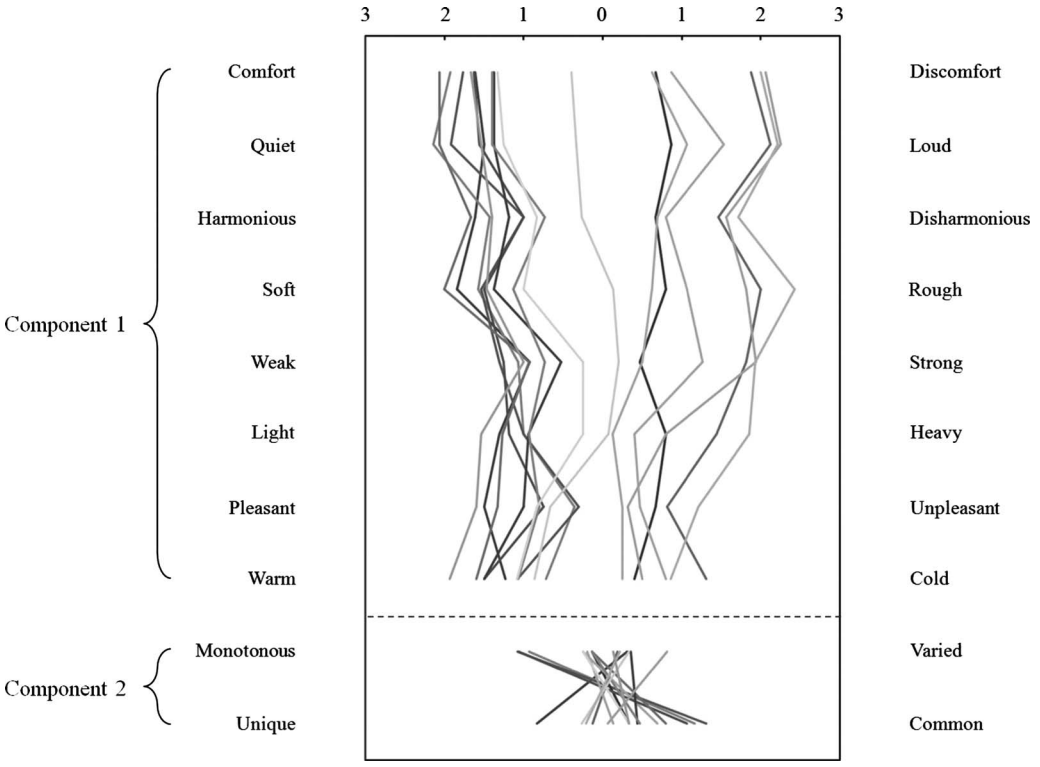


FIG. 3. Semantic differentials for the perception of urban soundscape, where each line corresponds to the mean rating for each area.

the percentage of responses that select exceeding "5" in the 11-point scale. From the obtained exposure-response relationships, the fraction of (highly) annoyed persons can be estimated based on the noise exposure. 50% of persons were highly annoyed when the short-term noise exposure was around 69–73 dB in urban spaces, and were annoyed in the situation with around 66–67 dB.

The differences of the %HA curves for the 5-point verbal and the 11-point numerical scales were statistically significant ($p<0.01$). This indicates that annoyance responses would be different according to the evaluation scale used. Therefore, the evaluation scale should be chosen appropriately in accordance with the purpose of the survey and the participants. In this study, most of the participants responded that the 11-point numerical scale was much easier to evaluate the annoyance than the 5-point verbal scale. In contrast to the results of the %HA, the %A curves from the two scales were not statistically different.

2. Qualitative analysis results

The results of the semantic differential test of urban soundscape using adjectives are shown in Fig. 3. These ten adjectives were grouped into two components through factor analysis. Component 1 included eight adjectives: comfort, quiet, harmonious, soft, weak, light, pleasant, and warm. Only two of the adjectives (monotonous and unique) were categorized into component 2. Components 1 and 2 explained 79% and 15% of the total variance, respectively. Component 1 is related to the comfort, loudness, and pitch sensations, and component 2 represents the temporal variation and the characteristics that are comparable with common urban soundscapes. The variation in subject scores for component 1 was much greater than that of component 2. This indicates that the adjectives of component 1 are more useful to characterize the soundscape than those of component 2. It was expected that the uniqueness of each site can be evaluated by a pair of adjective, "unique-common," but the differences between sites were not large in terms of uniqueness. This is because the nonacoustical components as well as dominant noise source of the sites were similar.

Table I presents the correlation coefficients between each component and the measured sound level, and between each component and the subjective evaluation results that were obtained from the field survey. Both component 1 and component 2 were correlated with the sound level and the subjective evaluation results, but component 1 had much higher correlation coefficients than component 2. Correlation coefficients between component 1 and the sound level, loudness, and annoyance were all positive values but the correlation coefficient between component 1 and the acoustic comfort was a negative value. Contrary to component 1, component 2 had negative correlation coefficients with the sound level, loudness, and annoyance, and a positive correlation coefficient with the acoustic comfort.

TABLE I. Correlation coefficients between each component and the measured sound level, and between each component and the subjective evaluation results that were obtained from the field survey.

	L_{Aeq}	Loudness	Annoyance	Acoustic comfort
Component 1	0.73[a]	0.93[a]	0.91[a]	−0.96[a]
Component 2	−0.52[b]	−0.59[b]	−0.59[b]	0.54[b]

[a]$p<0.01$.
[b]$p<0.05$.

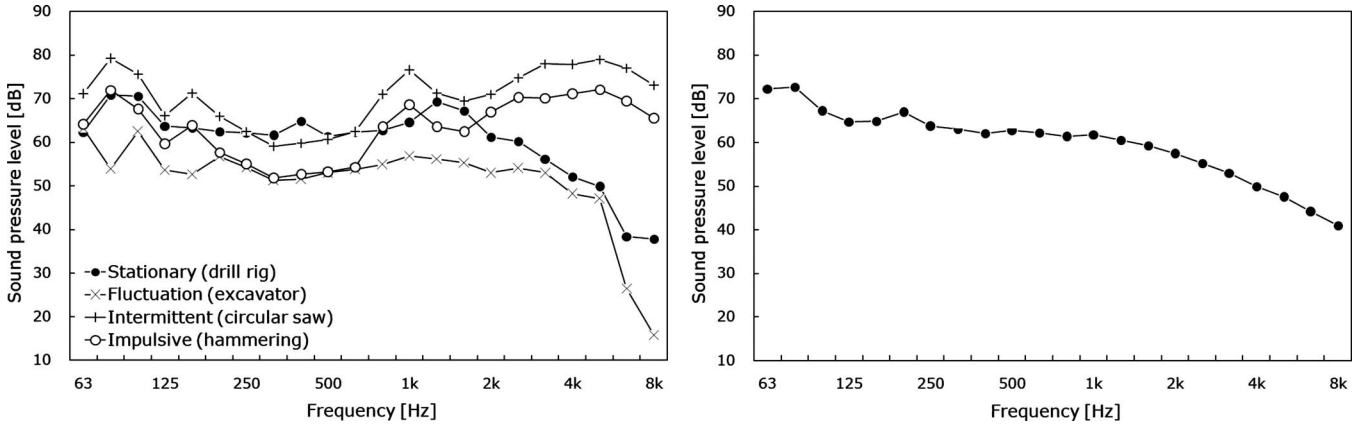


FIG. 4. Spectra of recordings: construction noises (left) and road traffic noise (right).

The difference between semantic scores of residential areas and those of open public spaces was not statistically significant. Perhaps this is because the dominant sound sources were road traffic and construction noises for nearly all the sites, so that the semantic scales used in this study were not sufficient to distinguish the differences between each soundscape. Another reason is that the subjects knew that road traffic and construction noises were the focus in the questionnaire, even though the semantic differentials concerned the appreciation of the soundscape as a whole. Therefore, it would have been better to add more questions to the questionnaire concerning the perception of other sound sources, and further multicultural and multidisciplinary research is needed in order to account for the diversity of urban soundscapes.

III. LABORATORY EXPERIMENT: ANNOYANCE FROM COMBINED NOISE SOURCES

While the total annoyance level of urban spaces with combined noise sources was obtained by conducting field surveys, as described above, it is not sufficient to investigate the effect of individual noise on the total annoyance because the type of construction noise and the noise levels of individual noises could not be controlled. Therefore, an auditory experiment was conducted under laboratory conditions in order to examine the effects of individual noises on the total annoyance.

A. Methodology

1. Recording of stimuli

Noises were recorded from nine construction sites in Korea and the UK. The sound recordings were made binaurally by using the artificial head (Neumann KU100) and ear microphones (type 4101, B&K). Recorded construction noises were classified into four groups, namely, stationary, fluctuation, intermittent, and impulsive sound, according to the temporal characteristics used in a previous study[24] and in the ISO 1996-1.[25] The noises from power generator and drill rig were classified into stationary sound, and the noises from excavator and circular saw were classified into fluctuation and intermittent sound, respectively. Hammer and concrete break noises were categorized to impulsive sound. The duration of each recorded stimulus varied from 15 s to 5 min, and the noise level ranged from 55 to 93 dBA in terms of L_{Aeq}. Road traffic noise with small fluctuations in noise level was also recorded. The road width was 40 m, and the average vehicle speed was around 60 km/h. Frequency characteristics of the measured construction noises representative of stationary, fluctuation, intermittent, and impulsive types are shown in Fig. 4, along with that of the measured road traffic noise.

2. Experimental design

The auditory experiments consisted of two parts, namely, the evaluation of individual noises and the evaluation of the combined noise sources. In the first part, the annoyance to individual noises including road traffic and the four types of construction noise was evaluated. The duration of the individual noises was 30 s, and their L_{Aeq} varied from 40 to 90 dBA, with an interval of 5 dB. The levels were adjusted in amplitude without the spectral adjustment to the stimuli, resulting similar spectral characteristics of the stimuli at different levels. A total of 55 stimuli for four types of construction noises and road traffic noise were presented to the subjects and the duration of the test for individual noises was around 30 min for each subject. The outline of the auditory experiment for the combined noise sources is shown in Table II. The independent variables were (1) the types of construction noise (stationary, fluctuation, intermittent, and impulsive), (2) the level of the construction noises and road traffic noise, and (3) the number of construction noise events. The number of events was expressed as the number of construction noises during a 1-min period. The L_{Aeq} of road traffic noise were fixed at 55 and 75 dBA corresponding to the noise exposure of most urban spaces.[26,27] Another reason for

TABLE II. Outline of the auditory experiment for combined noise sources, where the levels of the stimuli were adjusted according to the L_{Aeq} for 1 min.

Road traffic (dB)	Construction (stationary, fluctuation, intermittent, and impulsive) (dB)	No. of events (fluctuation, intermittent, and impulsive construction noise)
55	40, 50, 60, 70	2, 4, 6
75	60, 70, 80, 90	2, 4, 6

using this range was that approximately 40% of the EU population is exposed to road traffic noise with L_{Aeq} exceeding 55 dBA in the daytime, and 20% are exposed to levels exceeding 65 dBA.[28] The number of events for nonstationary construction noise (fluctuation, intermittent, and impulsive) was controlled to occur two, four, or six times, and the duration of each construction noise event was 10 s, while the duration of road traffic noise was 1 min. The levels of the stimuli were adjusted as a L_{Aeq} for 1 min regardless of the number of events. All 80 stimuli were randomly assigned to eight sessions and the duration of the test for combined noise sources was around 90 min for each subject.

3. Procedure of subjective evaluation

Subjects rated their level of annoyance when the individual noise or combined noise sources were presented to them. They were informed about the purpose of the experiments and were instructed about how to use the 11-point numerical scale. They were asked to respond to the following question when a binaural signal was played back to them through headphones (Sennheiser HD 600): *Which stimuli would be more annoying if you were exposed to it in the urban spaces?* The stimuli were presented to the subjects in the exact same order, and construction and road traffic noises were presented simultaneously. A visual image of an urban space was shown to the subjects before the auditory tests began.

Twenty subjects aged 24–30 participated in the experiment. Before the experiment, the hearing threshold level of the participants was tested with the use of an audiometer (Rion AA-77). The results showed that all of the participants had normal hearing. The experiments were conducted in a testing booth (4×3 m^2), where the background noise level was approximately 25 dBA.

B. Results

1. Individual noise sources

Annoyance ratings (with 0 as "not at all" and 10 as "extremely") for individual noises are plotted in Fig. 5 as a function of L_{Aeq}. The annoyance for both road traffic (T) and construction noises (C) increased with the increase in the noise level, as expected. The correlation coefficients between L_{Aeq} and the annoyance ratings were 0.96 for road traffic ($p < 0.01$) and 0.96–0.98 for the construction noises ($p < 0.01$). A significant difference between the annoyance ratings of the five stimuli was only found between the road traffic noise and the stationary construction noise ($p < 0.05$). The former was rated slightly higher than the latter. This result is somewhat different from the results of a previous study,[29] which showed that road traffic noise was rated lower than metal construction noise, perhaps because the metal construction noise included various noises such as hammering, banging, drilling and grinding.

2. Combined noise sources

The total annoyance ratings for stationary construction noise combined with road traffic noise are plotted in Fig. 6, where two cases are considered, with L_{Aeq} of road traffic

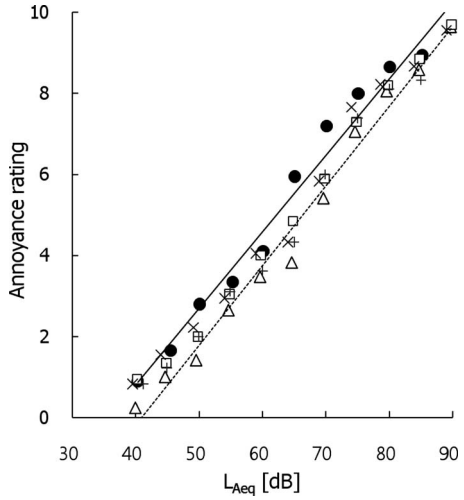


FIG. 5. Annoyance ratings for individual noises; road traffic noise and construction noises: (●) road traffic noise, (□) stationary, (□) fluctuated, (+) intermittent, (×) impulsive construction noise, (—) regression lines for road traffic noise, and (…) regression lines for stationary construction noise.

noise fixed at 55 and 75 dBA, respectively. The linear regression function (see Fig. 5) for C (stationary) in isolation and the interpolated annoyance caused by T only at L_{Aeq} of 55 and 75 dBA are shown in Fig. 6 as references. The total annoyance rating increased with the increase in the L_{Aeq} of construction noise. When the L_{Aeq} of road traffic noise was 55 dBA, the total annoyance ratings for the combined noise sources were significantly higher than the ratings for the more annoying sound in the combination when the L_{Aeq} of construction noise were 50, 60, and 70 dBA, respectively ($p < 0.05$). For the combination with construction noise at an L_{Aeq} of 60 dBA and road traffic noise at an L_{Aeq} of 55 dBA, expressed below as C60+T55 for the sake of convenience, the total annoyance ratings was considerably higher than the ratings for the more annoying sound (C60). But the difference between the annoyance ratings for the combined noise sources and the road traffic noise was not statistically signifi-

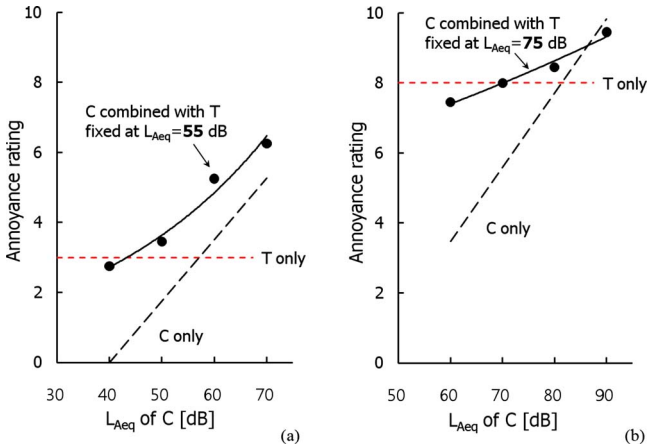


FIG. 6. (Color online) (a) Total annoyance for C (stationary) combined with T at an L_{eq} of 55 dBA. (b) Total annoyance for C (stationary) combined with T at an L_{eq} of 75 dBA. The solid lines represent the regression lines of the total annoyance.

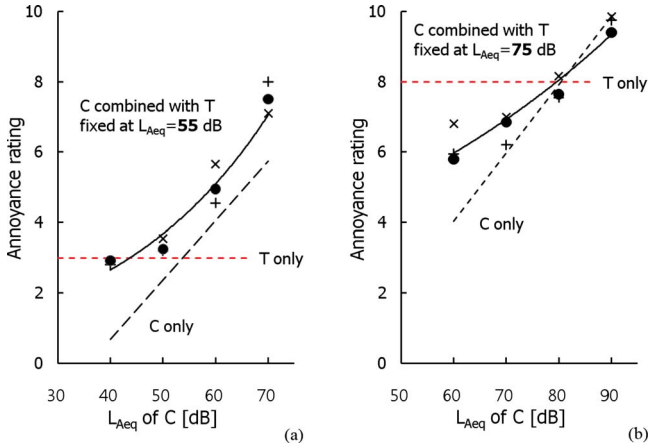


FIG. 7. (Color online) (a) Total annoyance for C (impulsive) combined with T at an L_{eq} of 55 dBA. (b) Total annoyance for C (impulsive) combined with T at an L_{eq} of 75 dBA [(●) two times, (+) four times, and (×) six times]. The solid lines represent the regression lines of the total annoyance.

cant in the case of C40+T55 because the L_{Aeq} for the combined noise sources (C40+T55) and the road traffic noise (T55) was almost the same.

When the L_{Aeq} of road traffic noise was 75 dBA, the total annoyance ratings for the combined noise sources were higher than the ratings for the more annoying sound in the combination only when the L_{Aeq} of the construction noise was 80 dBA ($p<0.05$). Under other conditions, the total annoyance for C60+T75 was rated lower than the more annoying sound, T75, ($p<0.05$). This is because that the construction noise with lower levels than road traffic noise affected total annoyance when the subjects can hear the construction noise. A previous study also showed that the ratings for combined annoyance were often lower than the annoyance caused by one of the sources alone.[11] The differences between the annoyance ratings for C70+T75 and the more annoying sound (T75), and C90+T75 and the more annoying sound (C90) were not significant.

The total annoyance ratings for the three nonstationary construction noises (fluctuated, intermittent, and impulsive) combined with road traffic noise are also obtained. The results of nonstationary construction noises showed a similar tendency; thus only the total annoyance ratings for the impulsive sound is plotted in Fig. 7, where two cases are again considered, with L_{Aeq} of road traffic noise fixed at 55 and 75 dBA. The number of nonstationary construction noise events varied and occurred two, four, or six times. The regression line for the construction noise was obtained by using the mean annoyance ratings for different numbers of events. The total annoyance ratings with the road traffic noise of 55 dBA were similar to the results of stationary construction noise. In the combinations of C50+T55, C60+T55, and C70+T55, the total annoyance ratings for the combined noise sources were significantly higher than the ratings of the more annoying sounds (T55, C60, and C70, respectively, $p<0.05$). But the difference between total annoyance ratings for C40+T55 and the more annoying sound (T55) was not statistically significant.

However, the total annoyance ratings with a higher road traffic noise of 75 dBA were somewhat different from the results of the stationary construction noise, suggesting the effects of temporal characteristics of the construction noise. In the combinations of C60+T75 and C70+T75, combined noise sources were rated significantly lower than the more annoying sound, T75 ($p<0.01$). This is also because the subjects can hear the impulsive sounds with lower levels in the specific regions where the level difference between road traffic and construction noises slightly decreased, and the impulsive sounds affected the rating of combined noise source. In the cases of C60+T75 and C70+T75, the difference between ratings for combined noise sources and more annoying sound much increased compared with the stationary noise. This is because it was easier to distinguish the impulsive sounds from the combined noise sources than stationary sounds. The total annoyance ratings for C80+T75 and C90+T75 were not statistically different from the ratings for the more annoying sound (T75 and C90).

The total annoyance was expected to increase as the number of events increased for the three kinds of nonstationary construction noises.[30] However, it can be seen from Fig. 7 that this is not the case. This is perhaps because the period of the noise exposure and construction noise is not enough for the awareness of the number of events.

The results of the auditory experiment indicate that the total annoyance ratings were almost the same or higher than that of the more annoying sound when the road traffic noise level was 55 dBA regardless the types of construction noise and level differences between road traffic and construction noises. Therefore, the level of construction noise should be reduced for the decrement of annoyance from combined noise sources. However, the total annoyance ratings were lower than that of the more annoying sounds in the case of combined noise sources with road traffic noise of 75 dBA when the level of construction noise was lower than that of road traffic noise (−15 dB for stationary, and −15 and −5 dB for impulsive sound) Thus, it is more effective to reduce the level of road traffic noise in order to decrease the total annoyance. These results would be useful for establishing the noise policy in urban spaces with combined noise sources.

IV. ENHANCEMENT OF SOUNDSCAPE USING WATER SOUNDS

In Secs. II and III, annoyance responses were investigated by applying the combined noises of urban open public spaces with the concept of soundscapes. In this section, the possibilities of applying water sounds, which usually exist in open public spaces, are examined to clarify its effects on soundscape enhancement. In other words, Secs. II and III could be regarded as the investigation of the present situation of urban soundscapes, and this section would be for one of the possible methods to overcome the problems revealed.

Natural sounds were introduced in this study in order to enhance the urban soundscape with sound masking effect. The selection of sound masker for the road traffic and construction noises, dominant sound sources of urban spaces, was conducted through auditory experiments. Appropriate signal to noise ratios (SNRs) between the sound masker and the urban noises were also investigated. Finally, the effect of

a sound masker was validated with experiments that applied the maskers to the combined noise sources with various sound levels.

A. Selection of natural sounds as maskers

Nine stimuli that cover various kinds of natural sound were introduced to investigate whether any of them were effective as maskers of urban noises. They were selected from a commercial sound effects CD (Sony Pictures Sound Effects Series), which included sounds called "Waterfall," "Rainfall', "Stream," "Waves of lake," "Birds in forest," "Birds in port," "Insects," "Bell of church," and "Wind." Spectral characteristics of the natural sounds were not modified, but their levels were adjusted for the experiments. Road traffic and construction noises were again chosen as representatives of urban noise. The road traffic noise used in Sec. III was applied in this experiment, and construction noises including various construction machineries were recorded at one of the field survey sites.

An auditory experiment was designed to determine which natural sounds were able to enhance the urban soundscape. The paired comparison method was adopted to evaluate the sound preference. 36 pairs consisting of natural sounds and road traffic noise were prepared and another 36 pairs were also made with natural sounds and construction noise. A typical stimulus is that the sounds of Rainfall and road traffic noise were presented to the subject simultaneously during the first 7 s and then after 1 s of silence, the sounds of Stream and the road traffic noise were presented in the same manner. After the presentation of each pair, subjects were asked to respond to the following question: *Which stimuli do you prefer if you were exposed to it in the urban space*? The SPL of natural sounds were 58 and 62 dBA for road traffic and construction noises, respectively, by considering the actual sound level of these noises. Each pair of stimuli was presented in a random order separated by an interval of 3 s, and a visual image of the actual site was presented to the subjects before the auditory tests began. Twelve subjects aged 20–30 participated in this experiment, and the stimuli were presented through headphones (Sennheiser HD 600) in a sound proof chamber.

The results of this auditory experiment are plotted in Fig. 8, where a larger value indicates a higher preference. The 11 subjects who passed the consistency test showed consistent judgments within a 95% confidence interval. This test also indicated that there was a significant ($p<0.05$) agreement among the subjects. As shown in Fig. 8, Stream and Waves of lake sounds were highly preferred with both road traffic and construction noises, which can therefore be used as an effective natural sound to mask urban noises. The sounds of Rainfall, Birds in port, and Wind showed relatively lower scale values of preference.

B. Effect of SNR on the preference of water sounds

Even though specific water sounds were determined as preferable sounds in urban spaces, a substantial difference in sound level between the water sound (masker) and the urban noise (masker) was needed to clarify the effect of the sound masking. Therefore, appropriate signal to noise ratios between the natural sounds and the urban noises were investigated as the next step. The road traffic noise and construction noises used above were again applied, and the two kinds of water sound, Stream and Waves of lake, were used as masker. No spectral correction was applied to the sound stimuli.

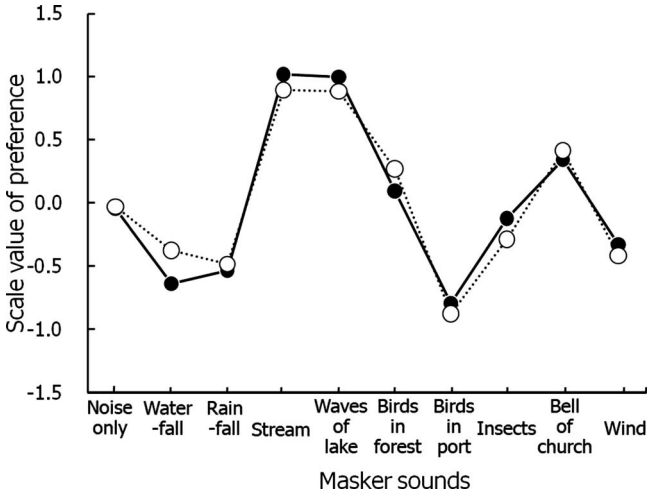


FIG. 8. Preferred natural sounds as masker of urban noises [(●) road traffic noise and (○) construction noise].

The SPL of the Stream and Waves of lake varied from 52 to 64 dBA in a step of 3 dB in the case of the road traffic noise of 58 dBA, and varied from 56 to 68 dBA in a step of 3 dB for the construction noise of 62 dBA in terms of L_{Aeq}. The paired comparison method was again employed in order to examine the effective level difference between the water sounds and the urban noises. Ten subjects aged 20–30 evaluated the sounds through headphones (Sennheiser HD 600) in a sound proof chamber.

The scale values of preference are shown in Fig. 9, where the nine subjects who passed the consistency test showed consistent judgments within a 95% confidence interval. This test also indicated that there was a significant ($p<0.05$) agreement among subjects. As shown in Fig. 9, the scale value was the highest when the relative SPL of Stream and Waves of lake was −3 dB compared to the road traffic noise. In the case of the construction noise, this value is −3 to 0 dB. It is interesting to note that the scale value decreased rapidly when the level of the water sound increased, perhaps

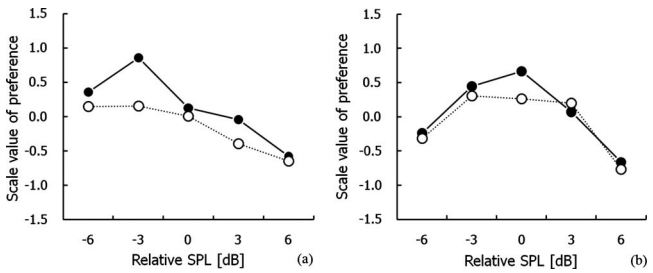


FIG. 9. (a) Scale value of preference according to relative presentation level of sounds of waters and road traffic noise. (b) Scale value of preference according to relative presentation level of sounds of waters and construction site noise [(●) Stream and (○) Waves of lake].

due to the increase in the total sound level. In other words, the results suggest that the water sounds are most effective to enhance the urban soundscape when the water sound is with the same level or 3 dB lower than the urban noises.

C. Effect of water sounds on combined noise sources

The effect of water sounds for the improvement of urban soundscape was examined by applying them to the combined urban noises, which reflects the auditory situation more realistically. An auditory experiment was performed in order to validate the results that were obtained in Secs. II and III using soundwalking and quantitative methodologies.

1. Recording and measurement of soundscapes

Two squares, one at the Town Hall and the other at the Railway Station, in Sheffield, UK, were chosen as the sites for the soundscape evaluations. The square in front of the Town Hall, which is known as the Peace Gardens, was enclosed by a two-lane motorway and several construction sites. Therefore, the Peace Gardens was suitable to investigate soundscape of urban spaces in terms of noise types. Aside from the urban noises, the Peace Gardens was also appropriate in examining the effects that water sounds have on urban noise. A fountain was located at the center of the square and there were several water features that were positioned at each of the entrances into the square. In the other site, the Sheaf Square in front of the Railway Station, there was a noise barrier integrated with a water stream and large fountains to mask the traffic noise from a busy road.

The sound recordings were made binaurally by using an artificial head (Neumann KU100) and a digital recorder (Fostex FR-2). Nine positions were chosen in the Peace Gardens for individual sound recordings. Urban road traffic noise (two positions), water sounds (four positions), babble (one position), and construction noise (two positions) were all recorded. The duration of the recordings was approximately 5 min at each position. In addition, video recordings were carried out along a path by using a video camcorder. The path was divided into three sections in order to experience the road traffic, water and babble, and construction noises, respectively. The total duration of the video recording was around 9 min, allowing 3 min for each section. Even though the video recording contained the audio track, only the video track was applied, and the sound recordings used in the stimuli were made by using the artificial head. This audiovisual stimulus was applied to the auditory tests prior to the main tests as a means to train the subjects. At the Railway Station Sheaf Square, ten positions were selected to record the road traffic noise (three positions) and water sounds (seven positions).

2. Auditory experiments

The road traffic noise and water sound recorded at the Sheaf Square, and the construction noise and babble noise recorded at the Peace Gardens were used as sound stimuli. The SPLs of each stimulus varied in the ranges of 76–85, 76–85, 66–75, and 78–87 dBA in a step of 3 dB for road traffic, water, babble, and construction noises, respectively.

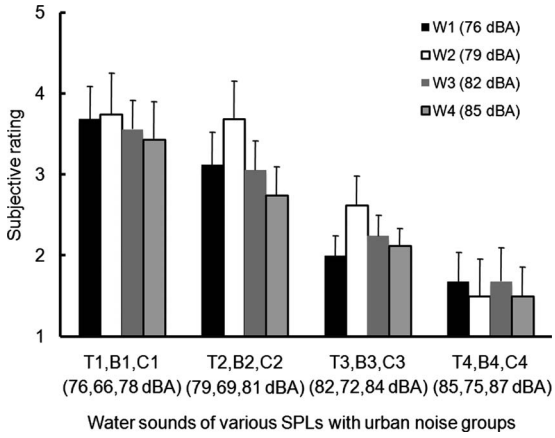


FIG. 10. Subjective rating of the combined noise sources and water sounds.

The levels of stimuli were determined by considering the actual sound levels in the case study sites. In total, 16 stimuli were used, where, for example, T1 indicates the stimulus of road traffic noise of 76 dBA, and W4 represents the water sound of 85 dBA. The duration of each stimulus was 3 s.

Auditory experiments were conducted under laboratory conditions for evaluating the effect of water sounds on combined noise sources of urban soundscapes. In order to simulate the combined noise sources, the road traffic, babble, and construction noises were combined as "T1, B1, C1," "T2, B2, C2," "T3, B3, C3," and "T4, B4, C4." Then, the water sounds with four levels were combined with the above four signals, such as "T1, B1, C1, W1," "T1, B1, C1, W2" to "T4, B4, C4, W3," and "T4, B4, C4, W4," forming 16 stimuli with combined noise sources. The overall SPLs of "T1, B1, C1," "T2, B2, C2," "T3, B3, C3," and "T4, B4, C4" were around 80.3, 83.3, 86.3, and 89.3 dBA, respectively; the SPL of the combinations with water sound was around 81.7 dBA, the lowest, in the case of T1, B1, C1, W1 and around 90.7 dBA, the highest, in the case of T4, B4, C4, W4. Each stimulus was presented to the 20 subjects through a headphone (AKG K-414P and Sennheiser HD 600). The subjects rated the preference to each combination of sounds by using a five-point scale, with 1 as not at all and 5 as extremely according to the following question: "Please, rate the sounds on your preference according to the 5-point scale."

3. Results and discussions

The preference ratings for urban noises combined with water sounds are plotted in Fig. 10. It can be seen that the preference for combined sources generally decreased with the increase in the noise level. At each combination of road traffic, babble, and construction noises, the effect of water sounds was different according to its SPL. Combined noise sources with a water sound of 79 dBA were evaluated as significantly higher than those with other water SPL in the cases of T2, B2, C2 and T3, B3, C3 ($p<0.01$). In other words, a water sound of 79 dBA was most effective in enhancing the urban soundscape when the SPL of the road traffic, babble, and construction noises varied from 79–82, 69–72, and 81–84 dBA in the combinations of T2, B2, C2 and T3, B3, C3, respectively. The effect of water sounds was

the highest in the combination of T2, B2, C2, where the level of water sounds was similar or slightly less than that of the dominant urban noises (road traffic and construction noises). This confirms the results of Sec. IV B that the urban soundscape can be enhanced when the level of the water sound is similar to or not less than 3 dB below the level of urban noises. However, the positive effects of the water sound on the urban soundscape became less when the overall level of the combined noise sources was around 90 dBA.

V. CONCLUSIONS

The total annoyance caused by combined noise sources in an urban space has been evaluated using field surveys and laboratory experiments. In order to reduce the annoyance of urban noises, the masking effects of various kinds of natural sounds have been examined, considering appropriate sound level. The study focused on the relationships between different sound sources without considering spectral variations in noises at different levels. The main conclusions are as follows.

- The synthesis curves are given for the relationship between the annoyance ratings for urban noises and short-term noise exposure. It was estimated that 50% of people were highly annoyed when the short-term noise exposure was around 70–73 dBA in urban spaces and were annoyed in the situation with around 66–67 dBA.
- From the results of semantic scales, it was revealed that the main factor was related to the acoustic comfort and loudness, and can be represented by such adjectives as "comfort," "quiet," and "weak."
- The annoyance ratings for construction noise in combination with road traffic noise were related to the types of construction noise (stationary and nonstationary) and different road traffic noise levels (55 and 75 dBA).
- The water sounds such as "Stream" and "Waves of lake" sounds were selected as an effective natural sound to mask urban noises. The level of the water sounds should be similar to or not less than 3 dB below the level of the urban noises.

[1]H. M. E. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[2]EU/DG Environment, "Directive 2002/49/EC of the European Parliament and the Council relating to the assessment and management of environmental noise," EU/DG Environment, Brussels, 2002.
[3]H. M. E. Miedema and H. Vos, "Noise annoyance from stationary sources: Relationships with exposure metric day-evening-night level (DENL) and their confidence intervals," J. Acoust. Soc. Am. **116**, 334–343 (2004).
[4]T. H. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[5]B. Berglund, Mi. Goldstein, and T. Lindvall, "Loudness (or annoyance) summation of combined Community noises," J. Acoust. Soc. Am. **70**, 1628–1634 (1981).
[6]B. Schulte-Fortkamp, "Noise from combined sources: How attitudes towards environment and sources influence the assessment," in Proceedings of the Internoise 99, Fort Lauderdale, FL, 1999.
[7]B. Schulte-Fortkamp, "The noise from combined noise sources—Are there any standards?," in Proceedings of the Noise-Con 2005, Minneapolis, MN (2005).
[8]H. V. C. Howart and M. J. Griffin, "Subjective response to combined noise and vibration: summation and interaction effect," J. Sound Vib. **143**, 443–454 (1990).
[9]S. M. Taylor, "A comparison of models to predict annoyance reactions to noise from mixed sources," J. Sound Vib. **81**, 123–138 (1982).
[10]C. G. Rice and K. Izumi, "Factors affecting the annoyance of combinations of noise sources," in Proceedings of the Institute of Acoustics 8 (1986).
[11]D. Botteldooren and A. Verkeyn, "Fuzzy models for accumulation of reported community noise annoyance from combined sources," J. Acoust. Soc. Am. **112**, 1496–1508 (2002).
[12]H. M. E. Miedema, "Relationship between exposure to multiple noise sources and noise annoyance," J. Acoust. Soc. Am. **116**, 949–957 (2004).
[13]International Standard Organization (ISO), "Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys," ISO/TS 15666 (2003).
[14]R. M. Schafer, *Our Sonic Environment and the Soundscape: The Tuning of the World* (Destiny Books, Vermont, 1994).
[15]B. Schulte-Fortkamp and A. Fiebig, "Soundscape analysis in a residential area: An evaluation of noise and people's mind," Acta Acust. **92**, 875–880 (2006).
[16]B. Berglund and M. E. Nilsson, "On a tool for measuring soundscape quality in urban residential areas," Acta Acust. **92**, 938–944 (2006).
[17]M. Raimbault, "Qualitative judgements of urban soundscapes: Questioning questionnaires and semantic scales," Acta Acust. **92**, 929–937 (2006).
[18]C. Semidor, "Listening to a city with the soundwalk method," Acta Acust. **92**, 959–964 (2006).
[19]E. Thompson, *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America 1900–1933* (MIT, London, 2002).
[20]D. Dubois, C. Guastavino, and M. Raimbault, "A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories," Acta Acust. **92**, 865–874 (2006).
[21]J. Kang, *Urban Sound Environment* (Taylor & Francis, New York, 2007).
[22]G. R. Watts, L. Ragonesi, K. V. Horoshekov, and R. J. Pheasant, "Investigation of water generated sounds for masking noise and improving tranquility," in Proceedings of the Internoise 08, Shanghai, China (2008).
[23]J. Y. Jeon, K. H. Kim, and T. Yano, "Standardized noise annoyance modifiers in Korean according to the ICBEN method," J. Acoust. Soc. Kr. **22**, 56–61 (2003).
[24]H. Yoshiaga, T. Yoshida, and A. Hayashi, "Analysis of individual machine noise in construction," in Proceedings of the Internoise 2004, Prague, Czech Republic (2004).
[25]International Standard Organization (ISO), "Acoustics—Description, measurement and assessment of environmental noise—Part 1: Basic quantities and assessment procedures," ISO/TS 1966 (2003).
[26]Road Facts 2000, An overview of Australian and New Zealand road systems, Report of Austroads, 2000.
[27]J. M. Barrigón Morillas, V. G. Escobar, J. A. M. Sierra, R. Vílchez-Gómez, and J. M. Vaquero, "A categorization method applied to the study of urban road traffic noise," J. Acoust. Soc. Am. **117**, 2844–2852 (2005).
[28]WHO, *Guidelines for Community Noise*, edited by B. Berglund, T. Lindvall, D. Schwela, and K. T. Goh (World Health Organization, Geneva, 2000).
[29]J. Vos, "Penalty for impulse noise, derived from annoyance ratings for impulse and road-traffic sounds," J. Acoust. Soc. Am. **77**, 193–201 (1985).
[30]J. M. Fields, "The effect of numbers of noise events on people's reactions to noise: An analysis of existing survey data," J. Acoust. Soc. Am. **75**, 447–467 (1984).

Interaural coherence for noise bands: Waveforms and envelopes

Neil L. Aaronson[a] and William M. Hartmann
Department of Physics and Astronomy, Biomedical and Physical Sciences Building, Michigan State University, East Lansing, Michigan 48824

(Received 17 June 2008; revised 10 December 2009; accepted 14 December 2009)

This paper reports the results of experiments performed in an effort to find a formulaic relationship between the interaural waveform coherence of a band of noise γ_W and the interaural envelope coherence of the noise band γ_E. An interdependence described by $\gamma_E = \pi/4 + (1-\pi/4)(\gamma_W)^{2.1}$ is found. This relationship holds true both in a computer experiment and for binaural measurements made in two rooms using a KEMAR manikin. Room measurements are used to derive a measure of reliability for the formula. Ultimately, a user who knows the waveform coherence can predict the envelope coherence with a small degree of uncertainty.
 [DOI: 10.1121/1.3290991]

PACS number(s): 43.55.Cs, 43.55.Hy, 43.66.Pn [LMW] Pages: 1367–1372

I. INTRODUCTION

The waveform interaural coherence is defined as the maximum of the interaural cross-correlation function for values of lag within limits, e.g., ±1 or ±2 ms. In psychoacoustical studies, it is often measured in one-third octave bands, which approximately correspond to the critical bandwidth of the human auditory system (Glasberg and Moore, 1990). Measurements of the waveform coherence are relevant in certain aspects of acoustics, and waveform coherence is often quoted in studies of rooms. It has been useful, for instance, in determining minimum microphone spacing when taking room measurements, especially in reverberant rooms (Jacobsen and Roisin, 2000; Kuster, 2008), and predicting acoustical sound quality of concert halls (Hidaka *et al.*, 1995). It has increasingly become an important measure in studies of both human and animal hearing, with applications to a variety of binaural detection and discrimination tasks (Bernstein and Trahiotis, 2007; van de Par *et al.*, 2001), binaural unmasking (Bernstein and Trahiotis, 1992; Culling *et al.*, 2001; Durlach *et al.*, 1986), auditory motion tracking (Grantham and Wightman, 1979), and is a basic component of most binaural models, especially involving interaural time difference detection (Colburn and Durlach, 1978; Jeffress *et al.*, 1962; Shackleton *et al.*, 2005). Waveform coherence is related to the apparent auditory source width, wherein the apparent auditory source width increases as coherence decreases (Ando, 1998; Blauert and Lindemann, 1986). Recently, it has also been linked to loudness perception (Edmonds and Culling, 2009). In summary, waveform coherence is an important factor in matters regarding modeling the human auditory system, the design of various listening spaces, applications involving virtual audio systems, and measurement of acoustical systems. In these applications, the effect of coherence on the listener is of primary importance.

At high frequencies, the ear is insensitive to interaural differences in the fine structure of signals, but differences in the envelopes of the signals become important (McFadden and Pasanen, 1978; Trahiotis *et al.*, 2005). Coding of fine structure in the binaural system is lost at frequencies above 1.3 kHz (Zwislocki and Feldman, 1956), and so only acoustical envelopes are represented in binaural neural activity. Consequentially, the coherence of the envelopes of the signals in the left and right ears, the "envelope coherence," is a more interesting binaural measure than the coherence of the waveforms of those signals for high-frequency bands (van de Par and Kohlrausch, 1995). Envelope coherence has been linked, for example, to binaural unmasking at high frequencies (Bernstein and Trahiotis, 1992). Across different environments, one expects that these two forms of coherence—waveform coherence and envelope coherence—will be related. In an acoustically dry environment where there are relatively few reflections off room surfaces, one expects that both coherences will be near unity. In a highly reverberant environment, both will be small.

The goal of the present work is to explore the relationship between the waveform coherence and the envelope coherence. In the literature, reported coherences are waveform coherences, but a reader may need to know the envelope coherence present in those studies. It is easier to measure the waveform coherence than to measure the envelope coherence since no manipulation of the measured signals needs to be done prior to calculating the waveform coherence. By contrast, in order to calculate envelope coherence, it is necessary to first construct the envelope of the signals, which involves either careful low-pass filtering or taking the absolute value of the analytic Hilbert transforms of the signals. The ideal result of this work would be a formula by which one could calculate the envelope coherence having measured the waveform coherence. The present work first finds such a formula for noise bands through a computer simulation. It then reports measurements of the two forms of coherence in two very different acoustical environments, as measured with a KEMAR manikin. It is shown that the most plausible relationship between coherences in actual room environments

[a]Present address: Department of Natural Sciences and Mathematics, The Richard Stockton College of New Jersey, P.O. Box 195, Pomona, NJ 08240. Author to whom correspondence should be addressed. Electronic mail: neil.aaronson@stockton.edu

agrees with the formula that emerges from the computer simulation. Finally, the actual room measurements are used to derive a measure of reliability for the formula.

II. ELEMENTARY IDEAS

The waveform coherence γ_W is the maximum value of the cross-correlation function between a signal in the left ear $x_L(t)$ and a signal in the right $x_R(t)$. The cross-correlation function is

$$\mathrm{cc}(\tau) = \frac{1}{T_D\sqrt{P_L P_R}} \int_0^{T_D} dt\, x_R(t)\, x_L(t+\tau), \tag{1}$$

where τ is the lag, and P refers to the average power of a signal, averaged over the same span of time, 0 to T_D. In practice, the duration of integration T_D is hundreds of times longer than the longest lag τ.

Thus the cross-correlation function is a measure of the similarity of a signal in the right ear to a signal as it occurs in the left ear at a later time—a time that is later by τ. The coherence is the maximum value of the cross-correlation function over all allowed values of τ. Therefore, the definition of coherence must include a limit on τ. In this article that limit is taken to be ± 1 ms, consistent with standards in the literature, e.g., Beranek (1986), since this is approximately the largest naturally occurring interaural delay in free field.

The envelope coherence γ_E is defined in the same way as the waveform coherence except that the envelopes $E_L(t)$ and $E_R(t)$ replace the waveforms x in the equation for the cross-correlation:

$$\mathrm{cc}_E(\tau) = \frac{1}{T_D\sqrt{P_{E,L} P_{E,R}}} \int_0^{T_D} dt\, E_R(t)\, E_L(t+\tau). \tag{2}$$

The envelopes of the signals are calculated as the absolute value of the analytic signal, for example,

$$E_R(t) = |x_R(t) + \imath \mathcal{H}\{x_R(t)\}|, \tag{3}$$

where $\mathcal{H}$ represents the Hilbert transform. P is then replaced by P_E, the envelope power, for instance,

$$P_{E,L} = \frac{1}{T_D} \int_0^{T_D} dt\, {E_L}^2(t). \tag{4}$$

Some limits to these formulas can be easily derived. If the signals in the left and right ears are identical then the waveform coherence and envelope coherence are both equal to 1.0, the value of the cross-correlation function at $\tau=0$. A value of 1.0 is the largest that the cross-correlation can ever be. Therefore, the largest possible coherence is 1.0 or 100%. Also, if the signals are sine tones, the waveform coherence is 1.0 because there is always some value of the lag for which the tones in the left and right ears are identical so long as the period of the sine is shorter than twice the maximum lag. In our case, twice the maximum lag is 2 ms, and the waveform coherence cannot be limited by that maximum lag so long as the sine period is shorter than 2 ms (frequency greater than 500 Hz). The envelope coherence for a sine is always 1, whatever the frequency, because the envelope is a constant.

In the present letter, the signals of interest are bands of Gaussian noise, which were chosen as representative of a generic signal. The bandwidths are given by the auditory filter widths of Glasberg and Moore (1990), approximately 1/3-octave.

For noise bands, the waveform coherence and the envelope coherence have well defined limits. As noted above, if the noises are the same in both ears, both the waveform coherence and the envelope coherence are 1.0. In anechoic environments, where the signals arriving at the two ears are very similar, waveform coherences arbitrarily close to 1 can be observed. That limit of perfect coherence places a simple restriction on any formula designed to relate the envelope coherence to the waveform coherence—the limit of perfect correlation. If, on the other hand, the noises in the two ears are perfectly uncorrelated then the waveform coherence is 0 because epochs when the two channels have the same sign will be matched, on the average, by epochs when the two channels have opposite signs. In practice, such a case of ideally perfect incoherence can occur only if the two waveforms are deliberately made to be orthogonal. However, in highly reverberant environments the waveform coherence can become close to zero. Thus the perfectly uncorrelated limit can easily be approached in practice.

In the perfectly uncorrelated limit, the envelope coherence will not be zero because the envelope is never negative. In this limit, the integral in the cross-correlation is easy to calculate because the temporal average of the product of the envelopes is equal to the product of the temporal averages. van de Par and Kohlrausch (1998) noted that for Gaussian noise, the envelope is Rayleigh distributed. For a Rayleigh distribution, the average value is related to the rms value by a factor of $\sqrt{\pi}/2$. Therefore, the coherence is $\pi/4$ or 0.7854. This value sets a second restriction on the relationship between envelope and waveform coherences, namely, when the waveform coherence is 0, the expected value of the envelope coherence is $\pi/4$. Because of the absolute value in Eq. (3), finding an analytic relationship between waveform and envelope coherences is made difficult, or perhaps impossible.

In an unpublished memorandum of 2004, Bernstein reported a computer experiment to determine the relationship between interaural waveform coherence and interaural envelope coherence. He tested the form

$$\gamma_E = \frac{\pi}{4} + b(\gamma_W)^n, \tag{5}$$

where parameters b and n were unrestrained (n is not an integer). He found that this form provided a reasonable fit to the results of his numerical experiments with $b=0.2142$ and $n=2.2$. Making use of the limits of perfect coherence, we may refine Bernstein's equation somewhat to give

$$\gamma_E = \frac{\pi}{4} + \left(1 - \frac{\pi}{4}\right)(\gamma_W)^n. \tag{6}$$

This is now an equation with one free parameter, the power n. It should be noted that Bernstein's value, $b=0.2142$, is very close to the exact value $1-\pi/4=0.2146$.

It is important to note that across the possible range of coherences (0–1), the value of n has a rather subtle effect on γ_E in Eq. (6). Consider, for example, $n=2.0$ versus $n=2.2$; the greatest difference between these two functions occurs at $\gamma_W=0.62$, where the difference is 0.0075—less than 1%. Because Eq. (6) is somewhat insensitive in this respect to the exact value of n, all reported values of n given in this paper will be rounded to the nearest tenths, and uncertainties (in the form of standard deviations) in the value of n smaller than ± 0.05 will be ignored.

III. COMPUTER EXPERIMENT

This section describes computer experiments using bands of noise. The experiments computed the waveform coherence and the envelope coherence and plotted one against the other for thousands of trials. The plots were then fitted to a function of the form of Eq. (6). In these computer experiments, two Gaussian noises were admixed in different proportions to lead to noise waveforms x_L and x_R with differing interaural coherences. The noises were 262 143 samples in length, sampled at a rate of 97.7 kHz. A total of 1001 different admixture proportions were used ranging from 0% to 100% and leading to waveform coherences ranging from 1.0 to 0.0. The 1001 different noises were time-domain filtered using gammatone filters[1] centered on ISO one-third octave frequencies from 160 to 10 000 Hz. Because there are 19 such filters, our experiment consisted of 19 019 different computations. Plots of waveform and envelope coherences, measured from the maxima of waveform and envelope cross-correlations within the limits $-1\ \text{ms} \leq \tau \leq 1\ \text{ms}$, are shown in Fig. 1 for six representative frequency bands. For each of the 19 bands, a value of $n=2.1$ resulted in the best fit to the data. 95% confidence intervals about the average value of n within each band were all less than ± 0.05.

The spread of the data about the best-fitting curve, measured as root-mean-square error (RMSE), is generally greater at lower frequencies than at higher frequencies. RMSEs were calculated on the scale of coherence (from 0 to 1). The calculated RMSEs are as high as 0.020 in the 160-Hz band and decrease steadily to a minimum of 0.0045 for the 10 000-Hz band. This is likely a bandwidth effect. Since bandwidth is proportional to the center frequency of the filter band, the number of degrees of freedom is proportional to the center frequency. One expects then that the RMSE will be inversely proportional to the square root of the center frequency. The best fit power law is close to that rule, $\text{RMSE}=0.13\ f^{-0.36}$.

IV. ROOM EXPERIMENT

A. Methods

The computer experiments do not necessarily correspond to any real acoustical environment. Therefore, an experiment was performed with the goal of measuring waveform and envelope coherences through ears and comparing them as in the computer experiments. To accomplish this, maximum length signals of order 18 ($2^{18}-1=262\ 143$) were played at a sample rate of 97.7 kHz through a loudspeaker in two different rooms. This made for signals about 2.6 s in duration, longer than the broadband reverberation times of the rooms used in this experiment. Signals were recorded in the ears of a KEMAR and filtered into 19 different bands with a gammatone filter bank. Cross-correlations of the waveforms and of the envelopes were computed within each band.

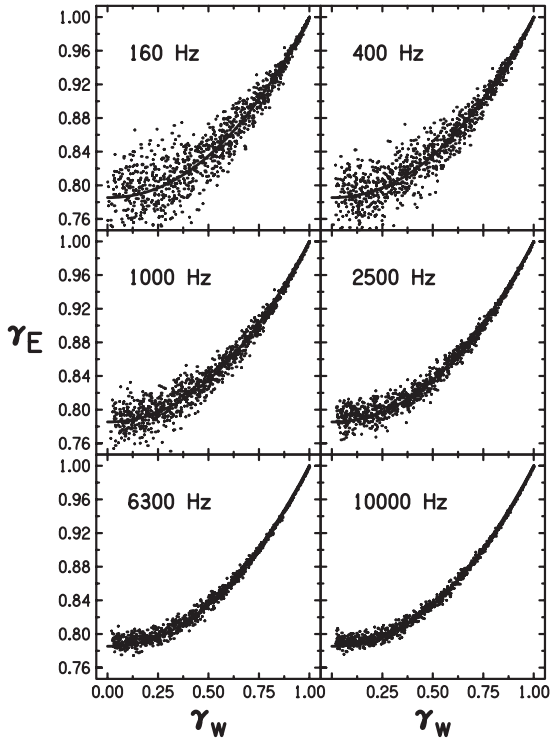


FIG. 1. Plots of envelope interaural coherence versus waveform interaural coherence for simulated binaural signals using Gaussian noise pairs in six different 1/3-octave bands. Each band contains 1001 coherence pairs, each of which is a single data point. A best-fitting line to Eq. (6) is shown as a thick solid line in each plot, though it may be obscured by the data points. The value of the power parameter n, to two significant figures, is 2.1 for each set of points in each panel.

One of the rooms used was an uncluttered laboratory space with hard surfaces and no carpeting, 6.5 m $\times$ 7.5 m $\times$ 4.5 m high. The broadband RT_{60} of this room was approximately 0.8 s; the RT_{60} in octave bands between 0.25 and 8 kHz were all between 0.7 and 0.9 s. The other room was a reverberant room, with dimensions 7.67 m $\times$ 6.35 m $\times$ 3.58 m high. The broadband RT_{60} of the reverberant room was approximately 2.0 s; the RT_{60} in octave bands was between 2.0 and 2.5 s for the bands from 0.5 to 2 kHz and was about 1.2 s in the 0.25- and 8-kHz bands. These rooms are, as reported in Hartmann *et al.*, 2005, rooms 10B and RR.

For all measurements, the KEMAR and loudspeaker were made to "face" one another. Measurements of coherence were made for four different distances between the loudspeaker and KEMAR—0.5, 1.0, 3.0, and 5.0 m. For each source distance, a measurement was made in ten different places in each room. For each measurement, both KEMAR and the loudspeaker were moved to different places in the room while keeping the distance between them the same. This yielded $19 \times 10 = 190$ waveform-envelope coherence pairs for each distance in each room.

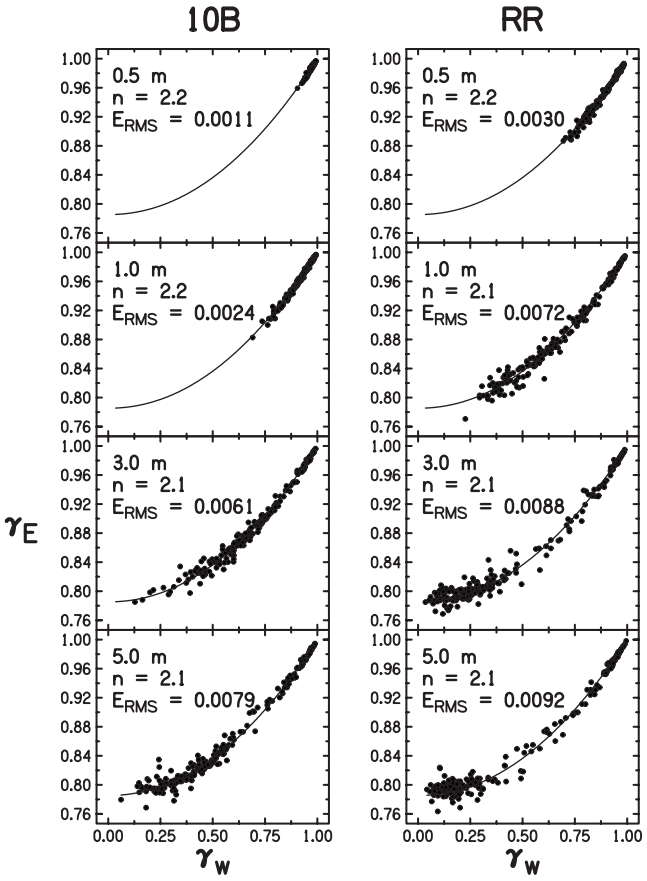


FIG. 2. Combined envelope versus waveform-coherence data across all 1/3-octave bands measured in rooms 10B and the reverberant room at each source distance. Trend lines show the best fit of the data in each panel to Eq. (6). The value of the fitting parameter n is shown in the upper-left corner of each panel. Because of the relative insensitivity of the curve to the exact value of n, only standard deviations greater than ± 0.05 are given. The rms error of the data to the best fit E_{rms} is given on the scale of coherence (0–1.0). Therefore, in room 10B at 1.0 m, where the mean γ_E is 0.96 and the rms error is 0.0024, the rms error is 0.25%.

B. Results

Within bands, Eq. (6) successfully describes the relationship between waveform and envelope coherences, though the value of the power parameter n within each band varies slightly. Consider, for example, room 10B at a distance of 3.0 m. The values of n found by fitting the data to Eq. (6) are as small as 1.9 in the 500-Hz band and as large as 2.2 in the 8000-Hz band. The variance in the best-fitting value of n is likely due to the narrow range of waveform coherences measured in individual bands. In some cases, where the total range of waveform coherences was small, the measured range of envelope coherences was also quite small. The regression which led to the value of n in those cases may be inaccurate. The average value of n across bands, plus and minus one standard deviation, was $n=2.1\pm0.1$.

The relative insensitivity of Eq. (6) makes determining an accurate value of n from a set of data with a small range of waveform coherences γ_W unreliable. Instead, data can be combined across all bands to get a picture of the waveform-envelope coherence relationship for any given room and source distance, as in Fig. 2. The resulting value of n given in Fig. 2 should apply reasonably well within any particular band. It can be seen that the combined data at every distance yield $n=2.1$ or 2.2. Standard deviations about these values were in all cases less than ±0.1. The collected coherences are lower on average as distance from the source increases and the direct-to-reverberant sound intensity ratio increases. Low-frequency bands still tend to have high coherence, and these form most of the points of high coherence in the cases where the distance was 3.0 or 5.0 m.

V. CONCLUSIONS

The relationship between envelope and waveform interaural coherences measured in real rooms agrees quite well with computer simulations which calculate coherences from Gaussian white noises, as computed in bands simulating auditory filters. The bands were approximately 1/3-octave in width. This close agreement suggests that the relationship between waveform and envelope coherences may be insensitive to the acoustic environment (reverberation time, source distance, etc.), though only two rooms and four distances were tested in this experiment. Within 1/3-octave bands, Eq. (6) provides a reliable relationship between waveform and envelope coherences, which can be expected to generalize to other rooms and distances. It may also be conjectured that Eq. (6) would generalize to bands of arbitrary width. Determination of the value of n measured within any given band is subject to error. By combining data across bands for a given source distance, a reliable value of n can be found that actually applies to any particular band. Current observations suggest that there may be little difference between rooms with vastly different reverberation times.

The exact value of n (i.e., a value with greater precision than mentioned here) is very sensitive to small variations in the data. This is because the exact value of n does not have a large impact on the shape of the curves, and so curves of similar shape can have quite different values of n. This also implies that the exact value of n may not be important to describe the behavior of the relationship between envelope and waveform coherences, but a "ballpark" estimate may suffice in most cases. In this study, a ballpark estimate is arrived at by rounding the relevant values of n to the nearest tenth. In both the computer experiment and in the rooms, the average value of n was 2.1. Thus, it is suggested that a reliable equation describing the relationship between waveform coherence γ_W and envelope coherence γ_E in any room and for any source distance is

$$\gamma_E = \frac{\pi}{4} + \left(1 - \frac{\pi}{4}\right)\gamma_W{}^{2.1}. \tag{7}$$

The envelope coherences arrived at using Eq. (7) are not exact. Especially for situations in which the value of the waveform coherence γ_W is small, the variance in the related envelope coherence γ_E seems significant (see Figs. 1 and 2). It would be useful to quantify an expected error in envelope coherence.

In order to quantify the expected deviation of envelope coherences from Eq. (7), waveform-coherence–envelope-coherence pairs were combined across all frequencies into one large set. Three such sets were constructed—one for

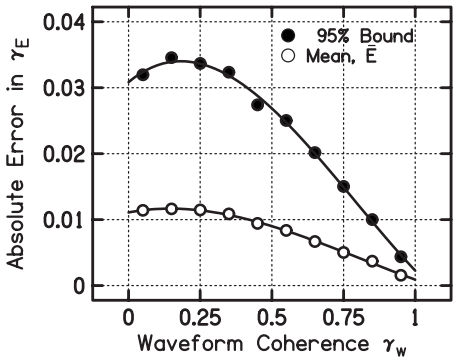


FIG. 3. The 95% bound curve (closed circles) gives the absolute error in envelope coherence for which 95% of the simulated coherences deviate by that amount or less from the coherence predicted by Eq. (7). The mean curve (open circles) gives the mean absolute error $\bar{E}$ relative to Eq. (7) for simulated envelope coherences. Both curves are third-order polynomials fitted to the absolute errors of the binned data.

each of the rooms and one for the computer simulation data. The data in each set were binned by the value of their waveform coherence into bins of width 0.1, thus creating ten bins across the range of waveform coherences. Within each bin, two methods of measuring error were employed: (1) calculation of the mean absolute difference $\bar{E}$ between the measured envelope coherences and envelope coherences as predicted by Eq. (7) and (2) calculation of the absolute difference from Eq. (7) for which 95% of the data deviate less than that, i.e., the "95% bound."

The absolute errors for the rooms turned out to be equal to or slightly less than the errors in the simulated coherences for both methods of error estimation. Thus, a conservative estimate of the absolute error in Eq. (7) is given by the errors measured in the computer simulation of coherences. The absolute errors for both the mean and 95% bound method are shown as a function of the waveform coherence γ_W in Fig. 3. This figure shows third-order polynomial function fits to the absolute errors. The mean absolute error $\bar{E}$ can then be described as a function of the waveform coherence by

$$\bar{E} = 0.011 + 0.0083\gamma_W - 0.031\gamma_W^2 + 0.013\gamma_W^3. \quad (8)$$

The 95% bound is approximately three times as large as the mean absolute error and reaches an upper limit less than 0.04. It is unlikely that listeners could discriminate between envelope coherences that are different by this amount. A user of Eq. (7) may expect the difference between envelope coherences calculated from Eq. (7) and the actual envelope coherences to be approximately $\pm\bar{E}$ on average and may expect envelope coherences to deviate no more than $\pm 3\bar{E}$ from Eq. (7). It should be noted that Eq. (7) works quite well, even for cases where the waveform coherence is low. At worst, the expected error is about ± 0.035.

It is interesting to wonder whether different waveforms would yield different results. An obvious choice to test would be speech signals. In a short study involving a measurement procedure identical to that used in the room experiment, male speech was used as a signal in place of broadband noise. The same distances between the speaker and KEMAR were used as in the above experiments, and measurements were made in both rooms, 10B and RR. Speech signals do not yield good signal-to-noise ratios at very high frequencies, so the analysis was limited to bands no higher than 10 kHz. Though an in-depth error analysis was not performed as it was for noise bands, it was found that $n=2.1$ works quite well to describe the relationship between waveform coherence and envelope coherence for these speech signals.

ACKNOWLEDGMENTS

The authors are grateful to Dr. L. R. Bernstein for sending them his memorandum and for useful conversations. Work was supported by NIDCD Grant No. DC00181.

[1]The gammatone filter is commonly used in auditory research because it provides a good fit to the impulse response of auditory nerve fibers as measured with the reversed correlation technique (Johannesma, 1972; Carney and Yin, 1988). The particular form used in our experiments was the form that appears on page 256 of Hartmann (1998) and used a filter order $\eta=4$ and Cambridge bandwidths (Glasberg and Moore, 1990).

Ando, Y. (**1998**). *Architectural Acoustics—Blending Sound Sources, Sound Fields, and Listeners* (Springer-Verlag, New York).
Beranek, L. L. (**1986**). *Acoustics* (American Institute of Physics, New York).
Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.
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–EL27.
Blauert, J., and Lindemann, W. (**1986**). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," J. Acoust. Soc. Am. **79**, 806–813.
Carney, L. H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.
Colburn, H. S., and Durlach, N. I. (**1978**). "Models of binaural interactions," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **IV**, pp. 467–518.
Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020–1029.
Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.
Edmonds, B. A., and Culling, J. F. (**2009**). "Interaural correlation and the binaural summation of loudness," J. Acoust. Soc. Am. **125**, 3865–3870.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.
Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer-Verlag, New York).
Hartmann, W. M., Rakerd, B., and Koller, A. (**2005**). "Binaural coherence in rooms," Acta Acust. **91**, 451–462.
Hidaka, T., Beranek, L. L., and Okano, T. (**1995**). "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," J. Acoust. Soc. Am. **98**, 988–1007.
Jacobsen, F., and Roisin, T. (**2000**). "The coherence of reverberant sound fields," J. Acoust. Soc. Am. **108**, 204–210.
Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1962**). "Effect of interaural correlation on the precision of centering a noise," J. Acoust. Soc. Am. **34**, 1122–1123.
Johannesma, P. I. M. (**1972**). "The pre-response stimulus ensemble of neurons in the cochlear nucleus," in Symposium on Hearing Theory, pp. 58–69.
Kuster, M. (**2008**). "Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity," J. Acoust. Soc. Am. **123**, 154–162.

McFadden, D., and Pasanen, E. G. (**1978**). "Binaural detection at high frequencies with time-delayed waveforms," J. Acoust. Soc. Am. **63**, 1120–1131.
Shackleton, T. M., Arnott, R. H., and Palmer, A. R. (**2005**). "Sensitivity to interaural correlation of single neurons in the inferior colliculus of guinea pigs," J. Assoc. Res. Otolaryngol. **6**, 244–259.
Trahiotis, C., Bernstein, L. R., Stern, R. M., and Buell, T. N. (**2005**). "Interaural correlation as the basis of a working model of binaural processing: An introduction," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 238–271.
van de Par, S., and Kohlrausch, A. (**1995**). "Analytical expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157–3169.
van de Par, S., and Kohlrausch, A. (**1998**). "Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research," J. Acoust. Soc. Am. **103**, 3605–3620.
van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2001**). "A consideration of the normalization that is typically included in correlation-based models of binaural detection," J. Acoust. Soc. Am. **109**, 830–833.
Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Speech enhancement using an equivalent source inverse filtering-based microphone array

Mingsian R. Bai,[a] Kur-Nan Hur, and Ying-Ting Liu
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

(Received 6 August 2009; revised 16 December 2009; accepted 20 December 2009)

This paper presents a microphone array technique aimed at enhancing speech quality in a reverberant environment. This technique is based on the central idea of single-input-multiple-output equivalent source inverse filtering (SIMO-ESIF). The inverse filters required by the time-domain processing in the technique serve two purposes: de-reverberation and noise reduction. The proposed approach could be useful in telecommunication applications such as automotive hands-free systems, where noise-corrupted speech signal generally needs to be enhanced. SIMO-ESIF can be further enhanced against uncertainties and perturbations by including an adaptive generalized side-lobe canceller. The system is implemented and validated experimentally in a car. As indicated by numerous performance measures, the proposed system proved effective in reducing noise in human speech without significantly compromising the speech quality. In addition, listening tests were conducted to assess the subjective performance of the proposed system, with results processed by using the analysis of variance and a *post hoc* Fisher's least significant difference (LSD) test to assess the pairwise difference between the noise reduction (NR) algorithms. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3291684]

PACS number(s): 43.60.Ac, 43.60.Dh, 43.60.Fg, 43.60.Np [EJS] Pages: 1373–1380

I. INTRODUCTION

Signal processing using microphone arrays has found applications in teleconferencing, telecommunication, speech recognition, speech enhancement, hearing aids, and so forth.[1] In these applications, how to effectively communicate in noisy or reverberant environments has been one of the pressing issues. Array techniques such as the well known delay-and-sum (DAS) beamformer do not function well in such environments due to multiple reflections.[2] To enhance interference rejection, a superdirective beamformer was suggested for its excellent spatial filtering capability.[3] Adaptive arrays provided useful alternatives for interference rejection.[1–13] The generalized side-lobe canceller (GSC) is an elegant approach in beamformer design in which a constrained optimization problem is converted into an unconstrained one from a linear algebraic perspective.[4] Its idea originated from linearly constrained minimum variance (LCMV) beamformer[5] and was first utilized by Owsley.[6] Griffiths and Jim[7] analyzed them and coined the term GSC. Their GSC corresponds to what we called in the present paper the Griffiths–Jim beamformer (GJBF). The key notion of GSC hinges at the use of the blocking matrix that further enhances the performance achievable by the fixed beamformer regardless of how it is implemented. However, oftentimes, adaptive implementation of GSC is preferred for two practical reasons. First, a fixed GSC requires the knowledge of covariance matrix, which is computationally prohibitive in real-time applications. Second, adaptive GSC is more robust in the presence of background noises, pointing errors, and other system uncertainties. There are various forms of adaptive implementation. Among them, GJBF is perhaps the most well known, which eliminates the need for on-line calculation of the signal correlation matrix and is robust against uncertainties and perturbations of the system. However, the GJBF still could fail in a reverberant environment in which multiple reflections cause problems of beamforming.[8] One example of such environment is the interior of a car cabin that is notoriously known to be a noisy and reverberant environment for speech communication. The speech signals tend to be corrupted by noises from the engine, tire, wind, etc., and reflections from the window, dashboard, seats, ceiling, etc.

In this paper, a microphone array technique is proposed for processing speech signals in noisy and reverberant environments. The idea of the technique originated from an equivalent source inverse filtering (ESIF) technique[9] developed for noise source identification purposes. However, the model in this paper is based on a single-input multiple-output (SIMO) structure, while the previous paper is based on a multiple-input multiple-output (MIMO) structure. This seemingly minor difference leads to many distinctive features in the implementation. First, unlike MIMO approach, the SIMO-ESIF formulation results in only one single focus, which simplifies tremendously the filter design that requires only simple phase conjugation and scaling, without having to design explicitly complicated inverse filters. The propagation matrices are basically represented by a vector $\mathbf{h}$, which renders the term $\mathbf{h}^H\mathbf{h}$ in the inverse filter a nonzero scalar and regularization is literarily unnecessary. Second, by posing the problem within a SIMO framework, inverse filters are designed based on measured acoustical plants, or systems to be controlled,[14] that include the effects of not only direct propa-

[a]Author to whom correspondence should be addressed. Electronic mail: msbai@mail.nctu.edu.tw

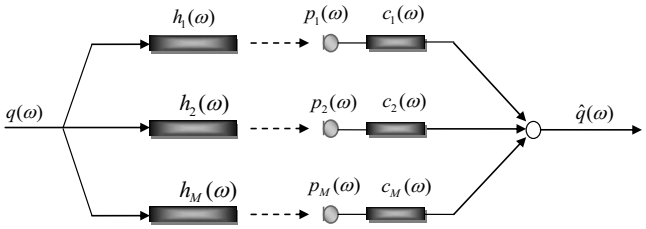


FIG. 1. The block diagram of the SIMO-ESIF algorithm. The parameter $q_m(\omega)$ is the input source, $h_m(\omega)$ is the propagation matrix, and $c_m(\omega)$ is the inverse filters.

gation from the source but also the reflections from the boundaries, which is different from the previous MIMO nearfield equivalence source imaging (NESI) that employs only free-field point source model. It follows that, with inverse filtering, both noise reduction and de-reverberation are fulfilled at the same time. Finally, the present work and the previous approach have totally different purposes in application. The ESIF algorithm aims at speech enhancement for telecommunication, whereas the previous NESI algorithm is intended for noise source identification (NSI) and sound field visualization. Another unique feature of the proposed technique is the use of the GSC to further enhance the performance of the proposed technique. An exact blocking matrix (BM) differing from those used in traditional beamformers is derived in this paper by solving a LCMV problem with two mutually orthogonal subspaces.[5] The leaky least-mean-squares (LMS) algorithms is exploited for adaptive filtering in the multiple-input canceller (MC).[15,16]

The proposed algorithms were implemented for enhancing speech communication quality in a car by using a multichannel data acquisition system. Objective tests were carried out to evaluate the algorithms.[1] In addition, listening tests were conducted to assess the subjective performance of the algorithms with data processed by the multivariate analysis of variance (MANOVA) (Ref. 17) and the least significant difference (Fisher's LSD) *post hoc* test.

II. EQUIVALENT SOURCE INVERSE FILTERING

The central idea of the proposed SIMO-ESIF algorithm is introduced in this section. In Fig. 1, M microphones are employed to pick up the sound emitting from a source positioned in the farfield. In the frequency domain, the sound pressure received at the microphones and the source signal can be related by a $M \times 1$ transfer matrix $\mathbf{H}$ as follows:

$$\mathbf{P} = \mathbf{H}q(\omega), \tag{1}$$

where $q(\omega)$ is the Fourier transform of a scalar source strength, $\mathbf{P}=[p_1(\omega)\cdots p_M(\omega)]^T$ is the pressure vector with T denoting matrix transpose, and $\mathbf{H}=[h_1(\omega)\cdots h_M(\omega)]^T$ is the $M \times 1$ propagation matrix. The aim here is to estimate the source signal $q(\omega)$ based on the pressure measurement $\mathbf{P}$ by using a set of inverse filters

$$\mathbf{C} = [c_1(\omega)\cdots c_M(\omega)]^T, \tag{2}$$

such that $\mathbf{C}^T\mathbf{H} \approx \mathbf{I}$ and therefore

$$\hat{q} = \mathbf{C}^T\mathbf{P} = \mathbf{C}^T\mathbf{H}q \approx q. \tag{3}$$

On the other hand, this problem can also be written in the context of the following least-squares optimization problem:

$$\min_q \|\mathbf{P} - \mathbf{H}q\|_2^2, \tag{4}$$

where $\|\ \|_2$ denotes vector 2-norm. This is an overdetermined problem whose least-squares solution is given by

$$\hat{q} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{P} = \frac{\mathbf{H}^H\mathbf{P}}{\|\mathbf{H}\|_2^2}, \tag{5}$$

where the superscript H denotes Hermitian transpose. Comparison of Eqs. (3) and (5) yields the following optimal inverse filters:

$$\mathbf{C}^T = \frac{\mathbf{H}^H}{\|\mathbf{H}\|_2^2}. \tag{6}$$

If the scalar $\|\mathbf{H}\|_2^2$ is omitted, the inverse filters above reduce to the "phase-conjugated" filters, or the "time-reversed" filters in the free-field context. Specifically, for a point source in the free field, it is straightforward to show that

$$\|\mathbf{H}\|_2^2 = \sum_{m=1}^{M} \frac{1}{r_m^2}, \tag{7}$$

where r_m is the distance between source and the mth microphone. Since $\|\mathbf{H}\|_2^2$ is a frequency-independent constant, the inverse filters and the time-reversed filters differ only by a constant. In a reverberant environment, these filters are different in general. Being able to incorporate the reverberant characteristics in the measured acoustical plant model represents an advantage of the proposed approach over conventional methods such as the DAS beamformer.

In real-time implementation, the inverse filters are converted to the time-domain finite-impulse-response (FIR) filters with the aid of inverse fast Fourier transform (IFFT) and circular shift. Thus, the source signal can be recovered by filtering the pressure signals with the inverse filters $\mathbf{c}(k)$ as follows:

$$\hat{q}(k) = \mathbf{c}^T(k) * \mathbf{p}(k), \tag{8}$$

where k is discrete-time index, $\mathbf{c}(k)$ is the impulse response of the inverse filter, and " *" denotes convolution.

III. ADAPTIVE GSC-ENHANCED SIMO-ESIF ALGORITHM

The SIMO-ESIF algorithm can be further enhanced by introducing an adaptive GSC to the system. The benefit is twofold. The directivity of the array is increased by suppressing the interferences due to side-lobe leakage. Robustness of the array is improved in the face of uncertainties and perturbations. The block diagram of the GSC with M microphones is shown in Fig. 2. The system comprises a fixed beamformer (FBF), a MC, and a BM.[11] The FBF aims at forming a beam in the look direction so that the target signal is passed and signals at other directions are rejected. The $p_m(k)$ is the signal received at the mth microphones and $\hat{q}(k)$ is the output signal of the FBF at the time instant k. The MC consists of

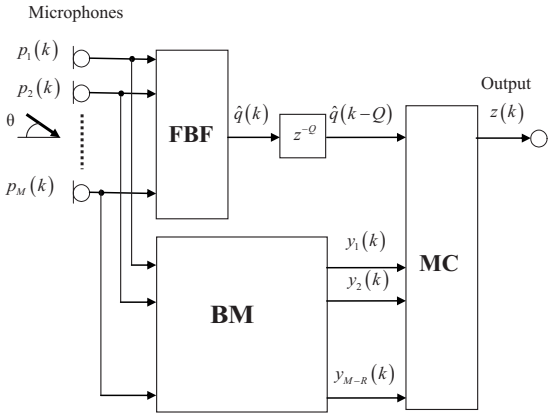


FIG. 2. The block diagram of the GSC, comprised of the FBF, the BM, and the MC.

multiple adaptive filters that generate replicas of components correlated with the interferences. The components correlated with the output signals $y_m(k)$ of the BM are subtracted from the delayed output signal $\hat{q}(k-Q)$ of the FBF, where Q is the number of modeling delay. Contrary to the FBF that produces a main-lobe, the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed though, hence the name "blocking matrix." The GSC subtracts the interferences that "leak" to the side-lobes in the subtractor output $z(k)$ and effectively improves spatial filtering.

A. Formulation of the blocking matrix

The purpose of the GSC depicted in Fig. 3 lies in minimizing the array output power, while maintaining the unity gain at the look direction (0-deg broadside is assumed here), which can be posed in the following constrained optimization formalism:[5]

$$\min_{\mathbf{w}} E\{|z|^2\} = \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{pp} \mathbf{w} \tag{9}$$

subject to

$$\mathbf{H}^H \mathbf{w} = 1, \tag{10}$$

where z is the array output signal, $\mathbf{R}_{pp}=E\{\mathbf{p}\mathbf{p}^H\}$ is the data correlation matrix, $E\{\}$ symbolizes the expected value, $\mathbf{H}$ is the frequency response vector corresponding to the propagation paths from the source to each microphone, and $\mathbf{w}$ is coefficient vector of the array filters. This constrained optimization problem can be converted into an unconstrained one by decomposing the optimal filter $\mathbf{w}$ into two linearly independent components belonging to two mutually orthogonal subspaces: the constraint range space $R(\mathbf{H})$ and the orthogonal null space $N(\mathbf{H}^H)$.

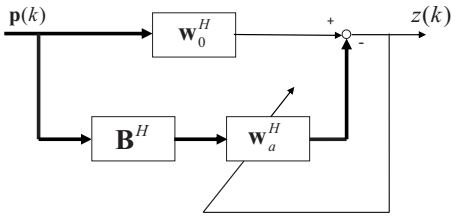


FIG. 3. The block diagram of the SIMO-ESIF-GSC algorithm. The parameter $\mathbf{p}(k)$ is the microphone signal, $\mathbf{B}^H$ is the BM, and $\mathbf{w}_0^H$ is the FBF.

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{v}, \tag{11}$$

where $\mathbf{w}_0 \in R(\mathbf{H})$ is a fixed filter and $\mathbf{v}=\mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H)$ with $\mathbf{w}_a$ being an adaptive filter. It follows that

$$\mathbf{H}^H \mathbf{w} = \mathbf{H}^H(\mathbf{w} - \mathbf{B}\mathbf{w}_a) = \mathbf{H}^H \mathbf{w}_0 - \mathbf{H}^H \mathbf{B}\mathbf{w}_a \approx 1. \tag{12}$$

The fixed filter $\mathbf{w}_0$ represents the quiescent component that guarantees the essential performance of beamforming. The filter design is off-line since it is independent of the data correlation matrix. It turns out that the minimization can then be carried out in the orthogonal subspace ($\mathbf{v}$) without impacting the constraint.

Traditionally, various *ad hoc* blocking matrices have been suggested. These matrices are based on the simple idea that, for free-field plane waves incident from the farfield broadside direction, $\mathbf{H}=[1\ \ 1\cdots 1]^H$. Since $\mathbf{H}^H\mathbf{B}=\mathbf{0}$, blocking is ensured if the columns of $\mathbf{B}$ sum up to zero; e.g., subtraction of signals of adjacent channels is a widely used approach. However, for a complex propagation matrix in a reverberant field, these *ad hoc* blocking matrices are inadequate. As a major distinction between the present approach and the conventional approaches, we shall derive an exact blocking matrix for a more general acoustical environment.

To fulfill the condition that $\mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H) \Leftrightarrow \mathbf{H}^H\mathbf{B}\mathbf{w}_a=0$, the columns of $\mathbf{B}$ must be constructed from the basis vectors of $N(\mathbf{H}^H)$ such that $\mathbf{H}^H\mathbf{B}=\mathbf{0}$.

$$\text{Let } \mathbf{H} = [a_1, a_2, \ldots, a_n]^H, \quad \mathbf{x} = [x_1, x_2, \ldots, x_n] \in N(\mathbf{H}^H)$$

$$\mathbf{H}^H\mathbf{x} = 0 \Rightarrow a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$$

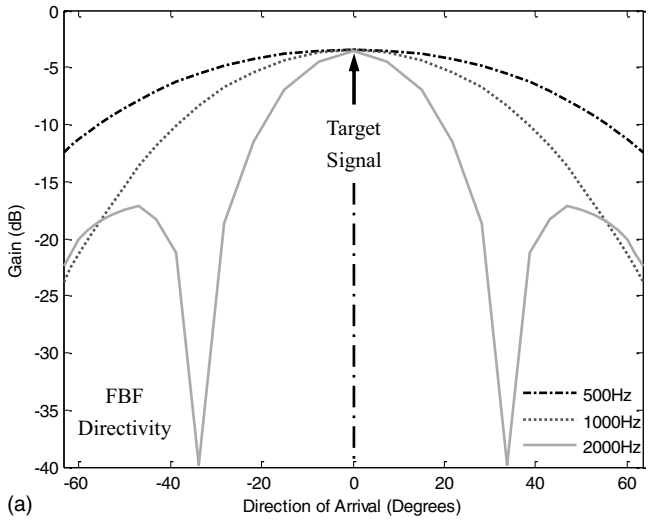
$$\text{If } a_1 \neq 0, \quad x_1 = -\frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3 - \cdots - \frac{a_n}{a_1}x_n$$

$$\text{Let } x_2 = \alpha_2, \quad x_3 = \alpha_3, \ldots, x_n = \alpha_n$$

$$\Rightarrow x_1 = -\frac{a_2}{a_1}\alpha_2 - \frac{a_3}{a_1}\alpha_3 - \cdots - \frac{a_n}{a_1}\alpha_n$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \alpha_2 \underbrace{\begin{bmatrix} -\frac{a_2}{a_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{v}_2} + \alpha_3 \underbrace{\begin{bmatrix} -\frac{a_3}{a_1} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{v}_3} + \cdots + \alpha_n \underbrace{\begin{bmatrix} -\frac{a_n}{a_1} \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_n}.$$

It is not difficult to see that $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ are linearly independent and form the basis of the null space $N(\mathbf{H}^H)$. Thus, the matrix $\mathbf{B}=(\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n)$ comprised of $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$ as its columns can be used as the blocking matrix; i.e.,



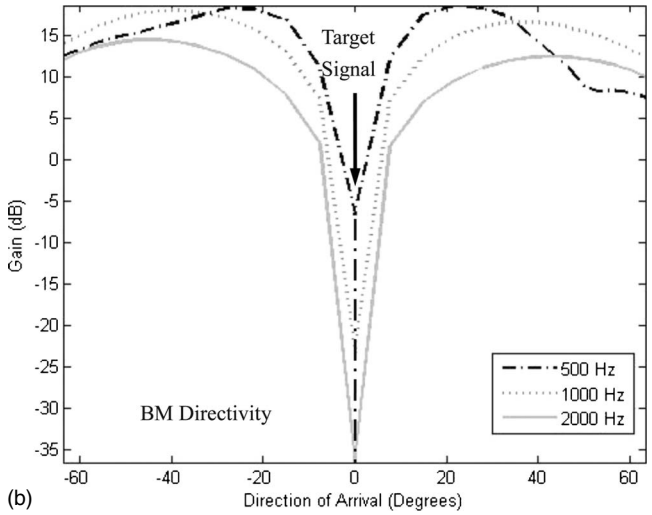


FIG. 4. The directivity pattern of the SIMO-ESIF-GSC algorithm at difference frequencies. (a) FBF with a main-lobe at the look direction. (b) BM with a null at the look direction.

$$\mathbf{B} = \begin{bmatrix} -\frac{a_2}{a_1} & -\frac{a_3}{a_1} & \cdots & -\frac{a_n}{a_1} \\ 1 & 0 & & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{13}$$

Physical insights can be gained by observing the beam patterns of the FBF and the BM shown in Fig. 4. Three sine wave signals at 500 Hz, 1 kHz, and 2 kHz are used to compare the performance of the BM between FBF, respectively. In the look direction, the FBF forms a main-lobe, whereas the BM forms a null so that the signal in the look direction is "blocked." The blocked path will attempt to further reduce the noise or interference outside the principal look direction (side-lobes). Note that the formulation above is in the frequency domain. For real-time implementation, the blocking matrix **B** needs to be converted to impulse responses using IFFT and circular shift.

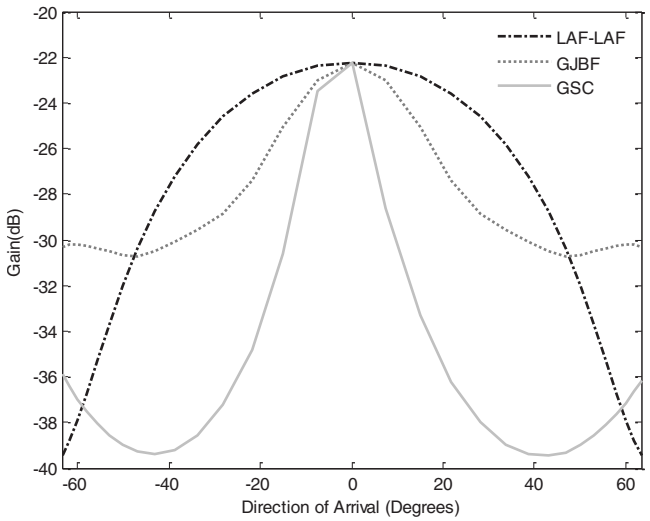


FIG. 5. The comparison of the beam patterns at 500 Hz obtained using the GJBF, LAF-LAF, and SIMO-ESIF-GSC algorithms.

B. Multiple-input canceller

In practice, the GSC is implemented using adaptive filters that are generally more robust than fixed filters. The need to compute the data correlation matrix $\mathbf{R}_{pp}$ is eliminated using such approach. For example, the leaky adaptive filters (LAFs) (Ref. 16) can be used in the MC block. LAFs subtract the components correlated with $y_n(k)$, $(n=0,\ldots,N)$ from $\hat{q}(k-Q)$, where Q is the modeling delay for causality. Let M_2 be the number of taps in each LAF and $\mathbf{w}_n(k)$ and $\mathbf{y}_n(k)$ be the coefficient vector and the signal vector of the nth LAF, respectively. The output of the MC module can be written as

$$z(k) = \hat{q}(k-Q) - \sum_{n=0}^{N-1} \mathbf{w}_n^T(k)\mathbf{y}_n(k), \tag{14}$$

$$\mathbf{w}_n(k) \triangleq [w_{n,0}(k), w_{n,l}(k), \ldots, w_{n,M_2-1}(k)]^T, \tag{15}$$

$$\mathbf{y}_n(k) \triangleq [y_n(k), y_n(k-1), \ldots, y_n(k-M_2+1)]^T. \tag{16}$$

The filter coefficients can be updated using the LMS algorithm

$$\mathbf{w}_n(k+1) = \mathbf{w}_n(k) + \mu z(k)\mathbf{y}_n(k), \tag{17}$$

where μ is the step size.

In Fig. 5, the beam pattern at 500 Hz of the proposed adaptive GSC algorithm is compared with two other conventional algorithms: GJB (Ref. 7) and LAF-LAF.[16] The GJBF algorithm adopts subtracted signals of adjacent channels as its BM block, whereas LAF-LAF algorithm uses adaptive filters to block the target signals. Both algorithms use adaptive algorithm identical to the MC block in Eq. (17). Clearly seen in Fig. 5, the proposed adaptive GSC algorithm attains the sharpest beam in the look direction with minimum side-lobes.

IV. ARRAY PERFORMANCE MEASURES

In the section, objective performance measures are defined for evaluating the array performance.[1] With the first microphone as the reference, the input signal to noise ratio (SNR) is defined as

$$\mathrm{SNR}_1(\mathrm{dB}) = 10 \log \frac{E\{x_1(k)^2\}}{E\{v_1(k)^2\}}, \tag{18}$$

where k is the discrete-time index, and $x_1(k)$ and $v_1(k)$ are the speech signal and the noise, respectively, received at microphone 1. The output SNR can also be defined for the array output

$$\mathrm{SNR}_A(\mathrm{dB}) = 10 \log \frac{E\{|\mathbf{c}(k)^T * \mathbf{x}(k)^2|\}}{E\{|\mathbf{c}(k)^T * \mathbf{v}(k)^2|\}}, \tag{19}$$

where $\mathbf{c}(k)$ is the impulse response of the inverse filter and " *" denotes convolution. Hence, the SNR gain is obtained by subtracting the output SNR from the input SNR.

$$\mathrm{SNR}(\mathrm{dB}) = \mathrm{SNR}_A - \mathrm{SNR}_1. \tag{20}$$

The SNRG quantifies the noise reduction performance due to array processing. However, noise reduction comes at the price of speech distortion in general. To assess speech distortion, a speech-distortion index (SDI) is defined as

$$\mathrm{SDI}(\mathrm{dB}) = 10 \log \frac{E\{x_1(k)^2\}}{E\{|x_1(k) - \mathbf{c}(k)^T * \mathbf{x}(k)|^2\}}. \tag{21}$$

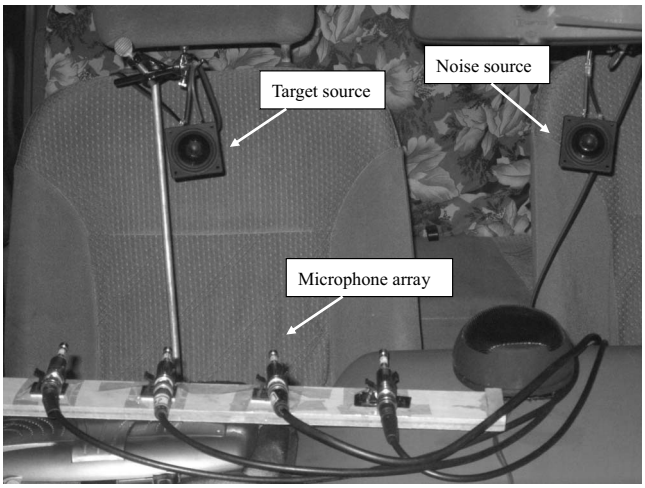
It is impractical to maximize both indices at the same time. The aim of array processing is then to reach the best compromise between the two indices.

V. OBJECTIVE AND SUBJECTIVE PERFORMANCE EVALUATIONS

The proposed algorithms have been examined experimentally in the vehicle compartment of a 2-l sedan. Figure 6 shows the experimental arrangement inside the car. Array signal processing algorithms are all implemented on National Instruments LABVIEW 8.6 and NI-PXI 8105 data acquisition system.[18] The sampling rate is 8 kHz. The sound pressure data were picked up by using a four-microphone (PCB 130D20) linear uniform array with inter-element spacing of 0.08 m. A loudspeaker positioned at (0.4 m, 0-deg) with respect to the array center was used to broadcast a clip of male speech in English, while another loudspeaker positioned at (0.3 m, 53-deg) was used to generate white noise as the interference.

Objective and subjective experiments were undertaken to evaluate the proposed methods. The SIMO-ESIF algorithm is used as the FBF and 512-tapped adaptive filters with step size μ=0.001 are used in the MC and LAF. There are variations of the SIMO-ESIF algorithm, depending on the plant model used and the filtering method in the FBF, as summarized in Table I. Two kinds of plant models, the free-field point source model and the measured plant in the car, are employed for designing the inverse filters. Two filtering methods, the inverse filtering and the time-reversed filtering,

(a)



(b)

FIG. 6. The experimental arrangement for validating the SIMO-ESIF algorithms. (a) The test car. (b) The experimental arrangement inside the car.

are employed in the FBF design. In addition, three variations of the processing methods with GSC are also included in Table I.

A. Objective evaluation

The objective measures SNR_1, SNR_A, SNRG, and SDI are employed to assess the performance of six proposed algorithms. The experimental results are summarized in Table II. By comparing the SIMO-ESIF and the SIMO-ESIF-GSC algorithms, the algorithms with GSC have attained signifi-

TABLE I. The acronyms and descriptions of six SIMO-ESIF algorithms.

Algorithm	Acronym	Description
SIMO-ESIF	PIF	Point source model-based inverse filtering
	MIF	Measured plant-based inverse filtering
	MTR	Measured plant-based time-reversed filtering
	GSC-PIF	Point source model-based inverse filtering
SIMO-ESIF-GSC	GSC-MIF	Measured plant-based inverse filtering
	GSC-MTR	Measured plant-based time-reversed filtering

TABLE II. The objective performance summary of the six algorithms.

	PIF		MIF		MTR	
SIMO-ESIF GSC	Without	With	Without	With	Without	With
SNR_I(dB)	3.79	3.79	3.79	3.79	3.79	3.79
SNR_A(dB)	12.96	15.28	15.56	19.19	13.58	13.66
SNRG(dB)	9.16	11.49	11.77	15.41	9.78	9.87
SDI(dB)	2.87	2.60	1.72	1.59	0.86	1.56

cantly higher noise reduction (SNRG) and lower speech distortion (SDI) than the algorithms without GSC. The time-reversed filters in general yield inferior performance than the inverse filters. The inverse filtering with the measured plant model considerably outperforms the point source model, e.g., SNRG of GSC-MIF=15.41 dB vs SNRG of GSC-PIF =11.49 dB. The implication of this result is that the inverse filters based on measured plant models have provided "dereverberation" effect in addition to noise reduction. Although the point source model-based inverse filtering (PIF) method tends to yield the least distortion, its noise reduction performance is also the worst. Figure 7 compares the time-domain wave forms obtained using SIMO-ESIF algorithm with and without GSC. Evidently, introduction of GSC has positive impact on noise reduction performance of the array.

Table III compares the proposed adaptive GSC algorithm and two other conventional algorithms, GJBF (Ref. 7) and CCAF.[11] The GJBF algorithm subtracts signals of adjacent channels as its BM block, whereas coefficient-constrained adaptive filtering (CCAF) algorithm uses constrained adaptive filters to block the target signals. Both algorithms use the adaptive algorithm identical to the MC block. The result revealed that the SIMO-ESIF algorithm augmented with the GSC outperformed the SIMO-ESIF algorithm without GSC. Among the GSC-based algorithms, the proposed GSC had attained the highest SNRG performance. The proposed GSC algorithm performed the best in noise reduction.

B. Subjective evaluation

Apart from the preceding objective tests, listening tests were conducted according to the ITU-R BS1116 (Ref. 19) to validate the algorithms. Subjective perception of the proposed algorithms was evaluated in terms of noise reduction and speech distortion. Specifically, three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were employed in the test. Fourteen participants in the listening tests were instructed with definitions of the subjective attributes and the procedures prior to the test. The subjective attributes are measured on an integer scale from 1 to 5. The participants were asked to respond in a questionnaire after listening. The six proposed algorithms previously used in the objective test are compared in the listening test. The test signals and conditions remain the same as in the preceding objective tests. A reference signal and an anchor signal are required in the ITU-R BS1116. In the test, the unprocessed signal received at the first microphone was used as the reference, while the lowpass-filtered reference was used as the hidden anchor. The mean and spread of the results of listening test are illustrated in Fig. 8. In order to access statistical significance of the results, the test data were processed using MANOVA (Ref. 17) with significance levels summarized in Table IV. Cases with significance levels below 0.05 indicate that statistically significant difference exists among the algorithms. In particular, the difference of the indices SIG and BAK

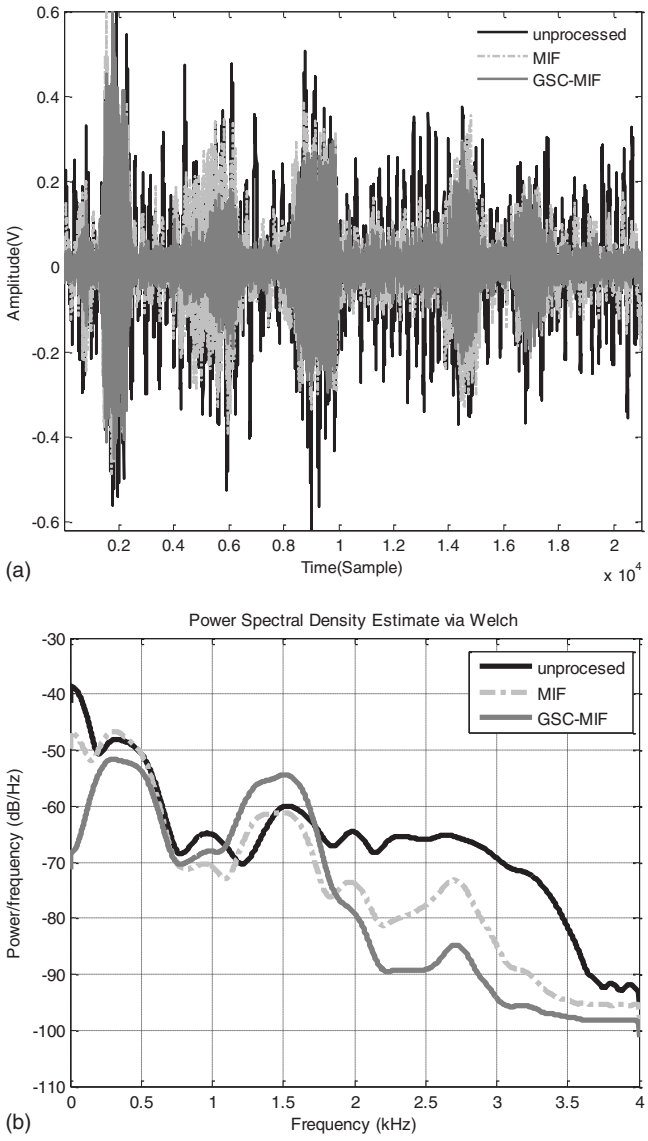


FIG. 7. The comparison of SIMO-ESIF-MIF algorithm and SIMO-ESIF-GSC-MIF algorithm by experimental measurement. (a) The time-domain wave forms. (b) The power spectral density functions.

TABLE III. The objective performance summary of the four beamforming algorithms including the ESIF, ESIF-GSC, GJBF, and CCAF algorithms.

	MIF			
Objective index	ESIF	ESIF-GSC	GJBF	CCAF
SNR_1(dB)	−1.04	−1.04	−1.04	−1.04
SNR_A(dB)	6.20	12.72	10.27	9.92
SNRG(dB)	7.24	13.76	11.31	10.96
SDI(dB)	1.86	1.42	2.49	1.90

among the six proposed methods was found to be statistically significant. Multiple regression analysis was applied to analyze the linear dependence of the OVL on the SIG and BAK. The regression model was found to be OVL=1.71+0.2 ×SIG+0.28×BAK. It revealed that the SIG has comparable but only slightly higher influence on the OVL than the BAK, whereas the indices SIG and the BAK are normally trade-offs. This explains why no significant difference can be found among methods in the OVL.

After the MANOVA, a *post hoc* Fisher's LSD test was employed to perform multiple paired comparisons. In Fig. 8, as opposed to the results of objective evaluation, the GSC-MIF algorithm performed not quite as expected in SIG. The price paid for the high noise reduction seems to be the signal distortion, which was noticed by many subjects. For the SIG index, the results of the *post hoc* test reveal that the GSC-PIF method outperforms the other methods. For the BAK index, the GSC-MIF method received the highest grade among all methods, which means that the inverse filtering approach has achieved both de-reverberation and noise reduction successfully. Despite the excellent performance in SIG, the PIF algorithm received low scores in BAK, which is consistent with the observation in the objective test. On the other hand, the GSC-PIF algorithm received higher SIG grade than plain PIF algorithm, indicating the GSC algorithm enhanced the SIMO-ESIF algorithm. However, the grades in the SIG and BAK indices showed no significant difference between the measured plant-based time-reversed filtering (MTR) and

TABLE IV. The MANOVA output of the listening test of the six algorithms. Cases with significance value p below 0.05 indicate that statistically significant difference exists among all methods.

	Significance value p		
Noise type	SIG	BAK	OVL
White noise	0.000	0.000	0.847

GSC-MTR algorithms. By comparing the BAK grade, all the proposed methods performed better than the reference signal.

Figure 9 compares the proposed GSC algorithm, GJBF,[6] and CCAF (Ref. 11) algorithms. The proposed GSC algorithm attained the highest BAK grades, while it also yielded lower SIG grades than the other algorithms. Apparently, the proposed GSC had attained the best performance in noise reduction at the expense of signal distortion. This is a typical trade-off for speech enhancement algorithms in general one has to face between signal distortion and noise reduction performance.

VI. CONCLUSIONS

A SIMO-ESIF microphone array technique has been developed for noisy automotive environments. Speech communication quality has been improved owing to the noise reduction and de-reverberation functions provided by the proposed system. With the use of specially derived BM of the GSC, the performance of SIMO-ESIF has been further enhanced.

The proposed algorithms have been validated via extensive objective and subjective tests. Overall, the results reveal that both de-reverberation and noise reduction can be achieved by using the SIMO-ESIF techniques. The methods exhibit different degrees in trading off noise reduction performance and speech-distortion quality. The MIF and GSC-MIF algorithms seem to have achieved a satisfactory compromise between these two attributes. All this leads to the conclusion that SIMO-ESIF-GSC-MIF proves effective in reducing noise and interference without markedly compromising speech quality.

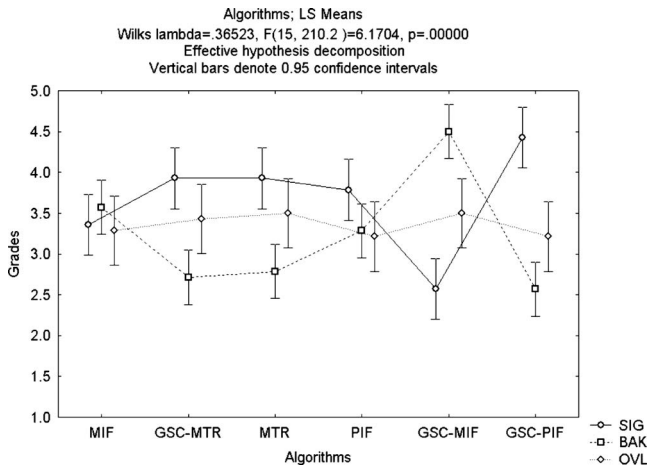


FIG. 8. The MANOVA output of the subjective listening test for the six SIMO-ESIF algorithms. Three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were evaluated in the test.

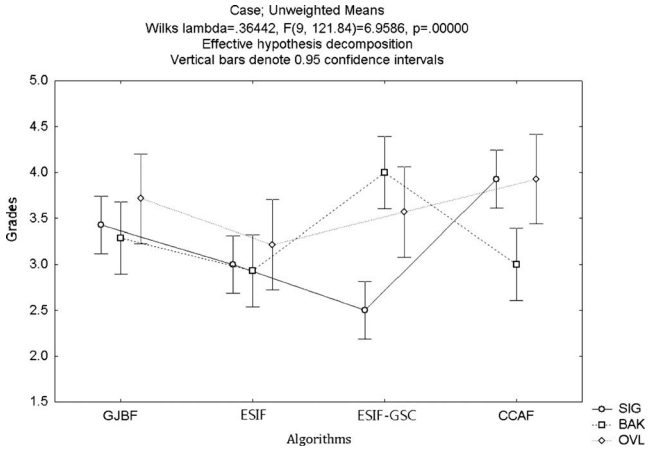


FIG. 9. The MANOVA output of the subjective listening test for the different GSC algorithms. Three subjective attributes including signal distortion (SIG), background intrusiveness (BAK), and overall quality (OVL) were evaluated in the test.

ACKNOWLEDGMENT

The work was supported by the National Science Council of Taiwan, Republic of China, under Project No. NSC 97-2221-E-009-010-MY3.

[1]J. Benesty, J. Chen, and Y. Huang, *Microphone Arrays Signal Processing* (Springer, New York, 2008).
[2]J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction techniques for hands-free speech recognition—A comparative study," in Robust Methods for Speech Recognition in Adverse Conditions (ROBUST99), Tampere, Finland (1999), pp. 171–174.
[3]J. Bitzer, K. D. Kammeyer, and K. U. Simmer, "An alternative implementation of the superdirective beamformer," Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October (1999).
[4]H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
[5]H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1365–1376 (1987).
[6]N. L. Owsley, "Source location with an adaptive antenna array," Technical Report No. AD0719896, Naval Underwater Systems Center, National Technical Information Service, Springfield, VA, 1971.
[7]L. J. Griffiths and C. W. Jim, "An alternative approach to linear constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34 (1982).
[8]J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multichannel noise reduction—Algorithms and theoretical limits," in Proceedings of the EURASIP European Signal Proceeding Conference (EUSIPCO), Rhodes, Greece, September (1998), Vol. **1**, pp. 105–108.
[9]M. R. Bai and J. H. Lin, "Source identification system based on the time-domain nearfield equivalence source imaging: Fundamental theory and implementation," J. Sound Vib. **307**, 202–225 (2007).
[10]M. Brandstein and D. Ward, *Microphone Arrays* (Springer, New York, 2001).
[11]O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone array with a blocking matrix using constrained adaptive filters," IEEE Trans. Signal Process. **47**, 2677–2684 (1999).
[12]Y. Grenier, "A microphone array for car environment," Speech Commun. **12**, 25–39 (1993).
[13]O. L. Frost III, "An algorithm for linearly-constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[14]G. F. Franklin, M. L. Workman, and D. Powell, *Feedback Control of Dynamic Systems*, 2nd ed. (Addison-Wesley, Boston, MA, 1993).
[15]I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," IEEE Trans. Antennas Propag. **40**, 1093–1096 (1992).
[16]O. Hoshuyama and A. Sugiyama, "A robust generalized sidelobe canceller with a blocking matrix using leaky adaptive filters," Electron Commun. Jpn. **80**, 56–65 (1997).
[17]S. Sharma, *Applied Multivariate Techniques* (Wiley, New York, 1996).
[18]National Instruments, http://sine.ni.com/nips/cds/view/p/lang/zht/nid/202630 (Last viewed 7/17/2009).
[19]ITU-R Rec. BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunications Union, Geneva, Switzerland, 1994.

A relationship between the far field diffraction pattern and the axial pressure radiating from a two-dimensional aperture[a)]

Edward H. Pees[b)]
Naval Undersea Warfare Center, 1176 Howell Street, Newport, Rhode Island 02841

(Received 16 June 2009; revised 16 December 2009; accepted 20 December 2009)

The diffraction of an acoustic wave by a two-dimensional aperture produces a sound field that can generally be represented at any point in space as a superposition of a continuum of plane waves. The mathematical formulation that facilitates this representation is known as the *angular spectrum of plane waves* method. The spectrum, in this representation, is a wavenumber spectrum obtained from a two-dimensional Fourier transform of the acoustic pressure (or velocity) distribution over the surface of the aperture boundary; a quantity which is also known to characterize the Fraunhofer diffraction pattern of the aperture. In this article, the angular spectrum method is used to formulate a mathematical relationship for two-dimensional apertures between the Fraunhofer diffraction pattern and a one-dimensional Fourier transform of the axial pressure. This relationship can be used to rapidly compute the axial pressure profile of the aperture if the boundary condition on the aperture is known and, in some cases, can be used as an inverse method. The approach is demonstrated for the cases of a flat circular piston and a flat rectangular piston undergoing harmonic motion in an infinite, rigid baffle. In the latter case, an analytical solution is also obtained.
 [DOI: 10.1121/1.3291685]

PACS number(s): 43.60.Gk, 43.30.Jx, 43.60.Pt, 43.20.El [EJS] Pages: 1381–1390

I. INTRODUCTION

The phenomenon of diffraction is the distortion of a wave field that cannot be attributed to either reflection or refraction, occurring when a wave, propagating in an otherwise homogeneous medium, encounters a disturbance in some property of the medium. It has long been known that when the undisturbed wave field takes the form of an infinite plane wave, this distortion is expressible as a transformation of the single plane wave into an infinite number of plane waves, each propagating in a unique direction.[1] This formulation, which has commonly come to be known as the *angular spectrum of plane waves* representation, has as its basis the early work of Weyl,[2] who in 1919 showed that the free space Green's function in three dimensions could be expressed as an expansion of plane waves, infinite in number and propagating in different directions. During the advent of modern computing and the development of the fast Fourier transform, the angular spectrum approach to solving diffraction problems gained additional popularity, since it is a computationally efficient Fourier transform-based method of propagating fields from one location to another. Though such applications began several decades ago,[3,1,4] the use of the angular spectrum method for investigating the characteristics of diffracting wave fields still continues among researchers in the fields of both acoustics and electromagnetism.[5–11]

In this article, we revisit the familiar problem of diffraction by a two-dimensional aperture in the context of the angular spectrum approach, for the purpose of demonstrating that the diffraction effect leads to the presence of a wavenumber spectrum, observable along the normal axis of the aperture, that can be directly related to the Hankel transform of the aperture function (the wavenumber spectrum in the plane of the aperture). As will be shown, this spectral relation provides a convenient means of computing the axial pressure profile of the aperture from values of the far field diffraction pattern and, in some cases, vice versa. Also, because the diffraction pattern has a direct relationship to the boundary condition on the aperture, it follows that knowledge of the axial field can potentially provide the means to quickly reconstruct a band-limited version of the boundary condition.

In Sec. II, we derive the theoretical relationship between the wavenumber spectrum of the acoustic pressure propagating along the axis of a rotationally symmetric, two-dimensional aperture and that of the normal velocity on the aperture boundary. In Sec. III, we utilize the widely recognized solution for diffraction by a flat circular piston in an infinite baffle to demonstrate computationally the validity of the results derived in Sec. II. Finally, in Sec. IV, we extend the formalism of Sec. II to address two-dimensional apertures of any shape, and utilize the result to derive a closed form solution for the axial pressure propagating from a rectangular aperture.

II. THEORY

We begin by considering monochromatic acoustic radiation, or single-frequency harmonic motion, for which the spatial distribution of the acoustic pressure field at any instant in time and any point $z>0$ can be represented in terms of an angular spectrum of plane waves as[12]

[a)]Portions of this work were presented in "Diffraction pattern of an axisymmetric aperture from axial measurements," 151st Meeting of the Acoustical Society of America, Providence, RI, June 2006.

[b)]Electronic mail: edward.pees@navy.mil

$$p(\mathbf{r}) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(k_x,k_y;z=0) \times e^{ik_x x}e^{ik_y y}e^{iz\sqrt{k_o^2-k_x^2-k_y^2}}dk_x dk_y, \quad (1)$$

where $P(k_x,k_y;z=0)$ is the two-dimensional spatial Fourier transform of the acoustic pressure $p(x,y,0)$ in the plane containing the aperture.

Given circular symmetry with respect to the z-axis, Eq. (1) can be more conveniently written in terms of a one-dimensional inverse Hankel transform as[13,14]

$$p(\rho,z) = \frac{1}{2\pi}\int_0^{\infty} P(k_\rho;z=0)J_0(k_\rho\rho)e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho, \quad (2)$$

where $\rho=\sqrt{x^2+y^2}$, and which, for points along the z-axis, simply reduces to

$$P(\rho=0,z)\frac{1}{2\pi}\int_0^{\infty} P(k_\rho;z=0)e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho. \quad (3)$$

We should note that $P(k_\rho;z=0)$ is no longer a two-dimensional Fourier transform of the acoustic pressure over the entire plane of the aperture, but the corresponding one-dimensional Hankel transform of the acoustic pressure along a radial line in the plane of the aperture.

From the linear inviscid force equation,[15] we have the relationship between the acoustic pressure and the normal component of the particle velocity

$$\frac{\partial p(\mathbf{r})}{\partial z} = i\omega_o\rho_o u_z(\mathbf{r}), \quad (4)$$

where ρ_o is the mass density of the medium, not to be confused with the spatial coordinate $\rho=\sqrt{x^2+y^2}$. In terms of Eq. (2), the partial derivative can be expressed as

$$\frac{\partial p(\rho,z)}{\partial z} = \frac{i}{2\pi}\int_0^{\infty}\sqrt{k_o^2-k_\rho^2}P(k_\rho;z=0) \times J_0(k_\rho\rho)e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho, \quad (5)$$

and, therefore, the normal velocity becomes

$$u_z(\rho,z) = \frac{1}{2\pi\omega_o\rho_o}\int_0^{\infty}\sqrt{k_o^2-k_\rho^2}P(k_\rho;z=0) \times J_0(k_\rho\rho)e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho. \quad (6)$$

Furthermore, since $u_z(\mathbf{r})$ satisfies the Helmholtz equation, it can also be expressed in terms of a plane wave spectrum[1]

$$u_z(\rho,z) = \frac{1}{2\pi}\int_0^{\infty} U(k_\rho;z=0)J_0(k_\rho\rho)e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho, \quad (7)$$

where $U(k_\rho;z=0)$ is the one-dimensional spatial Hankel transform of the normal velocity $u_z(\rho,z=0)$ along a radial line in the plane containing the aperture. Thus, comparing Eqs. (6) and (7), it is seen that

$$P(k_\rho;z=0) = \omega_o\rho_o\frac{U(k_\rho;z=0)}{\sqrt{k_o^2-k_\rho^2}}. \quad (8)$$

From this, we can then return to Eq. (3) and write the complex axial pressure for $z\geq 0$ as

$$P(\rho=0,z) = \frac{\omega_o\rho_o}{2\pi}\int_0^{\infty}\frac{U(k_\rho;z=0)}{\sqrt{k_o^2-k_\rho^2}}e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho, \quad (9)$$

which one may recognize as a form of King's integral[16] in the limit $\rho\to 0$.

Now that we have a relationship between the axial pressure and the normal velocity on the aperture boundary we will limit ourselves to values of the pressure field such that the upper limit of integration in Eq. (9) can be truncated at k_o, the cut-off wavenumber, beyond which the pressure field is evanescent,[17]

$$P(\rho=0,z) = \frac{\omega_o\rho_o}{2\pi}\int_0^{k_o}\frac{U(k_\rho;z=0)}{\sqrt{k_o^2-k_\rho^2}}e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho. \quad (10)$$

This approximation merely precludes the use of Eq. (10) for representing the *total* pressure field at values of z within a few wavelengths of the origin,[18] though it is still a valid representation for the propagating or *homogeneous* field.

Now, in the angular spectrum of plane wave representation, each plane wave component of the field has a wavevector magnitude of k_o; therefore, the magnitudes of the ρ- and z-components of a constituent wavevector always obey the constraint

$$k_o^2 = k_\rho^2 + k_z^2. \quad (11)$$

This constraint allows us to rewrite Eq. (10) in terms of an integration over k_z as

$$P(\rho=0,z) = \frac{\omega_o\rho_o}{2\pi}\int_0^{k_o} U(\sqrt{k_o^2-k_z^2};z=0)e^{ik_z z}dk_z. \quad (12)$$

The point to be made here is that Eq. (12) relates the axially observed, homogeneous pressure radiating from the aperture to the radially observed particle velocity in the plane containing the aperture, through a nonlinear mapping of a one-dimensional Hankel transform of the normal component of the particle velocity.

Note also that Eq. (12) takes the form of a band-limited inverse Fourier transform. Clearly this is equivalent to the proper inverse Fourier transform of a function that is identically zero outside the band of interest. We can therefore modify Eq. (12) to represent such a transform by the introduction of an appropriate windowing function.

That is, if $H(k_z)$ is the rectangular window defined as

$$H(k_z) \equiv 1 \quad \text{for} \quad 0\leq k_z\leq k_o;$$
$$\equiv 0 \quad \text{otherwise},$$

then multiplying this function with the integrand in Eq. (12) permits us to write the propagating part of the axial pressure field as

$$P(\rho=0,z)=\frac{\omega_o\rho_o}{2\pi}\int_{-\infty}^{\infty}U(\sqrt{k_o^2-k_z^2};z=0)H(k_z)e^{ik_zz}dk_z. \tag{13}$$

We are free to take the Fourier transform of Eq. (13), from which we obtain the rather simple result

$$P(k_z;\rho=0)=\omega_o\rho_oU(\sqrt{k_o^2-k_z^2};z=0) \quad \text{for} \quad 0\le k_z\le k_o. \tag{14}$$

Equation (14) is the principal result of this article. It reveals that the wavenumber spectrum of the acoustic pressure propagating along the axis of a rotationally symmetric, two-dimensional aperture can be obtained directly from that of the normal velocity on the boundary of such an aperture. Conversely, and under limited circumstances (discussed in the Appendix), the velocity spectrum and, by extension, the far field diffraction pattern can be obtained from knowledge of the axial pressure spectrum. This is clearly a relationship of some utility since, by Fourier inversion [i.e., Eq. (13)], the propagating part of the axial pressure associated with any two-dimensional, axisymmetric aperture can be directly computed if one can only specify the velocity boundary condition. Furthermore, from an analogous expression for the normal velocity, a band-limited version of a limited class of boundary conditions can be readily computed if one can specify the axial pressure. These assertions are demonstrated in Sec. III by considering the familiar problem of radiation from a flat, circular piston undergoing harmonic motion.

III. CIRCULAR PISTON

As an example, we now consider the problem of the flat, circular piston in an infinite rigid baffle, for which an analytical result for the axial pressure is widely available.[15,19,20] The far field directivity pattern for such an aperture is known to be given by[21]

$$D(\theta)=2\frac{J_1(k_oa\sin\theta)}{k_oa\sin\theta}. \tag{15}$$

We wish to demonstrate the validity of Eq. (14) by computing the radiation pattern of this aperture given only values of the axial pressure. Normalizing by $2\rho_oc u_o$, the axial pressure takes the form

$$p(\rho=0,z)=\frac{e^{ik_oz}-e^{ik_o\sqrt{z^2+a^2}}}{2}, \tag{16}$$

where a is the radius of the piston. Since an analytical result for the Fourier transform of Eq. (16) is not available, we compute a table of values and perform the Fourier transform numerically. Figure 1 shows a portion of the computed pressure values, and Fig. 2 shows the wavenumber spectrum obtained upon evaluation of the discrete, spatial Fourier transform.

As mentioned earlier, we are not considering evanescent energy in this problem; thus upon computing the Fourier transform of the axial pressure values, we truncate the resulting spectrum at the cut-off wavenumber k_o. What Fig. 2 effectively shows us then is the amplitude distribution of the plane waves comprising the angular spectrum of the diffracted pressure field, as a function of their propagation directions given by $k_z=k_o\cos\theta$; with θ being the angle measured from the normal axis of the aperture.

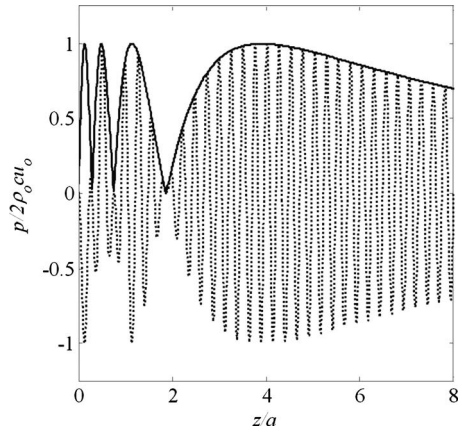


FIG. 1. Real part (dashed) and absolute value (solid) of complex axial pressure radiating from a baffled circular piston.

Applying the transformation given by Eq. (14) to these data yields the Hankel transform $U(k_\rho)$ of the aperture function, or aside from a multiplicative phase factor,[22] the Fraunhofer radiation pattern of the aperture. This is shown in Fig. 3, along with the theoretical result given by Eq. (15), with $k_\rho=k_o\sin\theta$.

As expected, the results are a precise match. One may notice, however, the uneven distribution of data points that occurs due to the nonlinear variable transformation given by Eq. (14). We further note that a band-limited approximation to the boundary condition can be recovered via an inverse Hankel transform of this quantity. Because of the uneven sample spacing, however, one must first interpolate these values to obtain a uniformly spaced set of data points to which a discrete Hankel transform can be applied. This could be considered a one-dimensional analog of the two-dimensional frequency domain interpolation technique associated with diffraction tomogramphy.[23] Figure 4 shows the

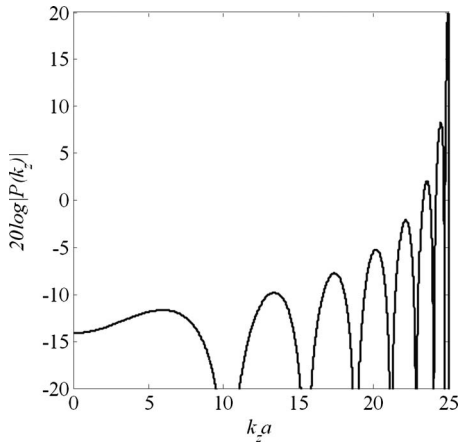


FIG. 2. Magnitude of the wavenumber spectrum obtained from a discrete Fourier transform of the axial pressure radiating from a baffled circular piston.

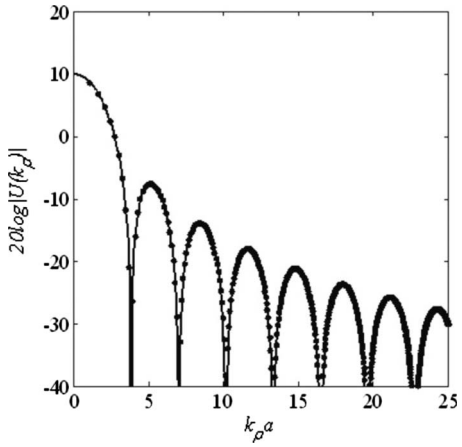


FIG. 3. Magnitude of the radial wavenumber spectrum (dots) of the acoustic velocity on the boundary of a baffled circular piston, obtained from the axial pressure spectrum via Eq. (14) vs the theoretical diffraction pattern (solid).

result of computing the inverse Hankel transform of $U(k_\rho)$ using linear interpolation to first obtain a uniformly sampled data set.

As a final demonstration, we show how the reconstruction process just described can be reversed to alternatively permit the computation of axial pressure fields from specified boundary conditions and their corresponding wavenumber spectra. That is, we show how the Fraunhofer diffraction pattern of an axisymmetric aperture can be used to compute the propagating part of the axial pressure radiating from the aperture. To see how this is achieved, let us assume that we are again dealing with the baffled circular piston having a Neumann boundary condition with a normal velocity amplitude that is unity on the piston face and zero on the baffle; we then have the well known Fraunhofer result for $U(k_\rho;z=0)$ (Refs. 13, 17, 19, and 24) [shown in Fig. 5(a) for $k_o a=25$] as follows:

$$U(k_\rho;z=0)=2\pi a\frac{J_1(k_\rho a)}{k_\rho}, \tag{17}$$

where a is the radius of the piston. Substituting this into Eq. (14) gives us an expression for the wavenumber spectrum of the axial pressure [see Fig. 5(b)]

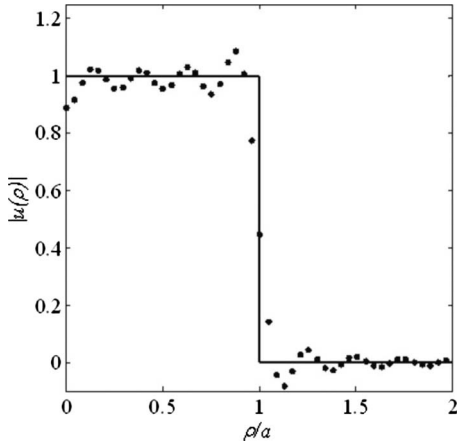


FIG. 4. Band-limited reconstruction (dots) of the velocity boundary condition at the surface of a baffled circular piston obtained from the axial pressure spectrum via a discrete inverse Hankel transform of Eq. (14). The solid line segments represent the ideal rectangular boundary condition.

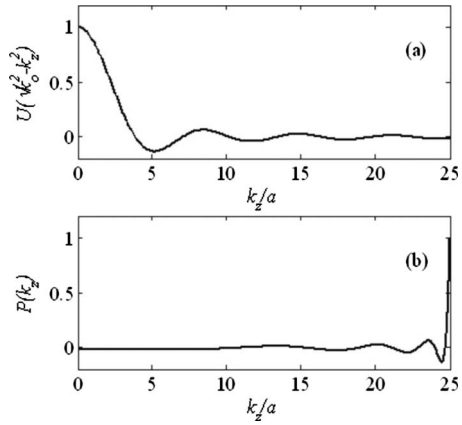


FIG. 5. Theoretical result for the Hankel transform of the ideal Neumann boundary condition associated with the baffled circular piston (a), and the corresponding axial wavenumber spectrum (b) obtained via Eq. (14).

$$P(k_z)=2\pi a\omega_o\rho_o\frac{J_1(a\sqrt{k_o^2-k_z^2})}{\sqrt{k_o^2-k_z^2}}. \tag{18}$$

We can see that by its very nature, the Fraunhofer diffraction pattern only provides that portion of the axial wavenumber spectrum up to the cut-off wavenumber. This is no surprise, of course, since the evanescent field, which corresponds to higher wavenumbers, is not observable in the Fraunhofer region. It implies, therefore, that the axial pressure obtained from a discrete Fourier transform of this wavenumber spectrum will have limited spatial resolution. To show this, we construct a discretely sampled spectrum using Eq. (18) for $k_o a=25$, and compute the axial pressure by performing a discrete inverse Fourier transform on the data. Figure 6 shows both the theoretical result from Eq. (16) and the result of numerically computing the inverse transform.

As can be seen, the resulting spatial resolution (which is $\Delta z/a=\pi/k_o a\cong 0.13$ in this case) is inadequate for capturing the lobe structure below $z/a\cong 2$. In an attempt to remedy this, our first inclination might be to simply zero-pad the wavenumber spectrum prior to Fourier inversion. However, as is evident in Fig. 7, the unwanted Gibbs phenomenon[25] arises due to the discontinuity that zero-padding introduces to the wavenumber spectrum at $k_z=0$.

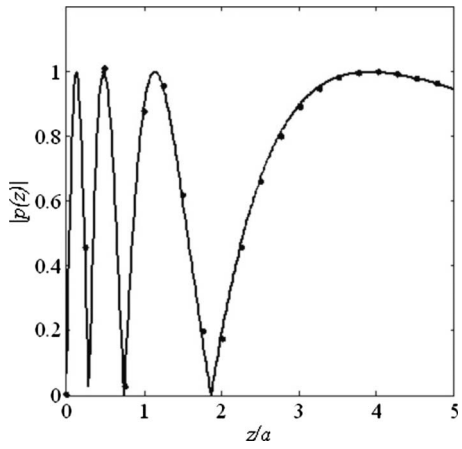


FIG. 6. Magnitude of the acoustic pressure propagating along the axis of a baffled circular piston, obtained from a discrete implementation of Eq. (12) (dots) vs the theoretical result given by Eq. (16) (solid).

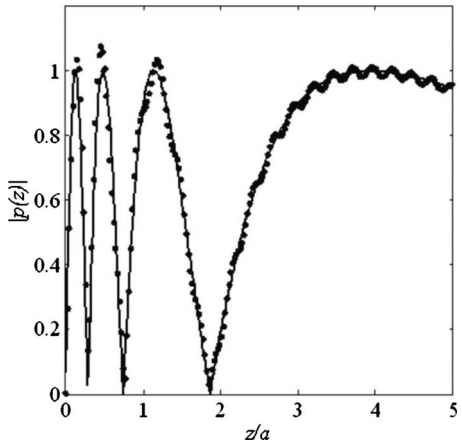


FIG. 7. Magnitude of the acoustic pressure propagating along the axis of a baffled circular piston, obtained from a discrete implementation of Eq. (12) with zero-padding to illustrate the effect of Gibbs phenomenon (dots) vs the theoretical result given by Eq. (16) (solid).

It is possible, however, to mitigate this effect somewhat if we consider the relative contributions of different portions of the wavenumber spectrum to the axial pressure. As noted earlier, the principal contribution comes from the plane wave component of the angular spectrum that propagates directly along the piston axis. Conversely, as the direction of propagation of the remaining plane wave components moves further from the axis of symmetry (i.e., the k_z-components of the wavevectors decrease), their contributions to the axial field generally diminish. This suggests that, if we can suppress the discontinuity at $k_z=0$ without affecting the spectrum at higher values of k_z, we may be able to recover a more precise representation of the axial pressure field. Such an objective is commonly met in signal processing applications with the use of shading functions.

To illustrate its effect, we apply a shading function to the pressure spectrum shown in Fig. 5 and recompute the inverse Fourier transform. The function used in this example is a 50% cosine-tapered (Tukey) window (see Fig. 8), and its effect on the axial pressure spectrum is illustrated in Fig. 9.

Performing the inverse Fourier transform on this shaded quantity yields the axial pressure profile shown in Fig. 10. As

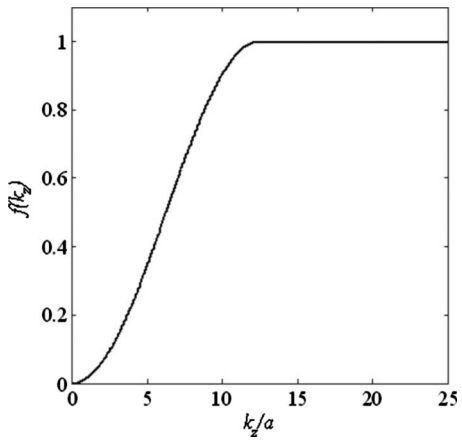


FIG. 8. Illustration of a 50% cosine-tapered (Tukey) shading function used in place of the ideal Heaviside function in Eq. (13) to mitigate the effects of Gibbs phenomenon observable in Fig. 7.

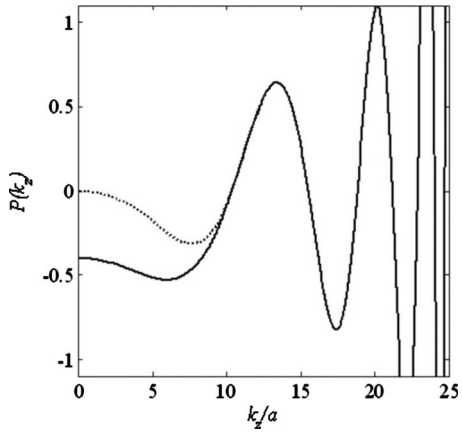


FIG. 9. Wavenumber spectrum of the axial pressure associated with a baffled circular piston, with (dashes) and without (solid) the application of the Tukey shading function shown in Fig. 8.

can be seen, the result is a successful elimination of the rippling observed in Fig. 7, along with, however, a loss of some information for small values of z/a. While this information loss is due, in part, to the absence of evanescent energy in our solution, it is also due to the additional restriction imposed on the bandwidth of the spectrum by the shading function, needed to obtain a well behaved quantity to which a discrete Fourier transform could successfully be applied.

Since the choice of shading function was somewhat arbitrary in this example, it is reasonable to expect that the accuracy of the computed pressure profile could be improved by developing a more optimal selection method. However, since our intent here is to merely demonstrate the validity of the formalism presented in Sec. II, we make no effort to do so in this article.

It may be informative though, to point out that while, in principle, Eq. (13) gives us a general result for the nonevanescent (propagating) part of the axial pressure; in practice, the Heaviside function $H(k_z)$ in the integrand should be

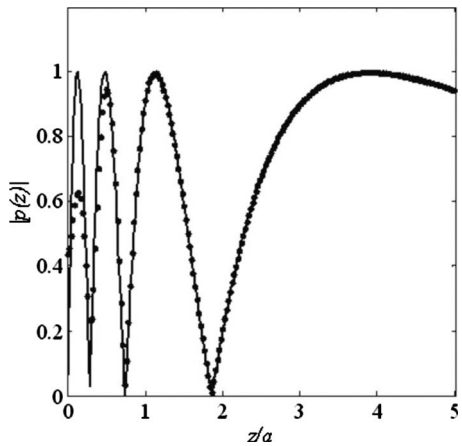


FIG. 10. Magnitude of the acoustic pressure propagating along the axis of a baffled circular piston, obtained from a discrete implementation of Eq. (13) (dots) with the Heaviside function replaced with a Tukey shading function. The departure between the discrete values and the theoretical curve (solid) occurs mostly due to the absence of evanescent energy in the formulation, but partially to the suppression by the Tukey function of some propagating modes as well.

modified to optimize the discrete implementation of this equation. In this sense, $H(k_z)$ is an ideal lowpass filter that suppresses only those spectral components above the cut-off wavenumber, while the shading function is a nonideal filter further suppressing some components of the spectrum below the cut-off wavenumber to permit a more accurate numerical reconstruction of the axial pressure.

IV. RECTANGULAR PISTON

Having now developed a technique for computing the axial pressure of a rotationally symmetric aperture from its Fraunhofer pattern (or the Hankel transform of the boundary condition), it will finally be shown how the same approach can be used to compute an "axial" pressure associated with any two-dimensional aperture for which the boundary condition (hence the far field diffraction pattern) is known. In the case of apertures that possess no particular symmetry the term axial becomes ambiguous, but will be retained nonetheless to facilitate the discussion. The baffled rectangular piston will be used as an example, wherein the technique will provide not only a computational result for the homogeneous axial pressure, but will also lead to an analytical solution.

Returning to Eq. (1), we write the axial pressure in Cartesian coordinates [rather than cylindrical coordinates, as in Eq. (3)] by setting $x=y=0$ as follows:

$$p(x=y=0,z\geq 0)=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P(k_x,k_y;z=0)\times e^{iz\sqrt{k_o^2-k_x^2-k_y^2}}dk_x dk_y. \tag{19}$$

We know from Eq. (8) that the relationship between the two-dimensional spatial Fourier transform of the pressure distribution in the plane at $z=0$ and that of the acoustic velocity normal to this plane can be expressed as

$$P(k_x,k_y;z=0)=\omega_o\rho_o\frac{U(k_x,k_y;z=0)}{\sqrt{k_o^2-k_x^2-k_y^2}}. \tag{20}$$

If we therefore substitute this expression into Eq. (19), we obtain, similar to Eq. (9), the acoustic pressure observable along the positive z-axis in terms of the velocity spectrum in the plane at $z=0$ as follows:

$$p(x=y=0,z\geq 0)=\frac{\omega_o\rho_o}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{U(k_x,k_y;z=0)}{\sqrt{k_o^2-k_x^2-k_y^2}}\times e^{iz\sqrt{k_o^2-k_x^2-k_y^2}}dk_x dk_y. \tag{21}$$

Now, even though we assume no symmetries of any kind, way may still rewrite this equation in cylindrical coordinates as

$$p(\rho=0,z\geq 0)=\frac{\omega_o\rho_o}{(2\pi)^2}\int_{0}^{\infty}\int_{0}^{2\pi}\frac{U(k_\rho,\varphi;z=0)}{\sqrt{k_o^2-k_\rho^2}}\times e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho d\varphi. \tag{22}$$

Or, limiting the integration over k_ρ to address only propagating waves,

$$p(\rho=0,z\geq 0)=\frac{\omega_o\rho_o}{(2\pi)^2}\int_{0}^{k_o}\int_{0}^{2\pi}\frac{U(k_\rho,\varphi;z=0)}{\sqrt{k_o^2-k_\rho^2}}\times e^{iz\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho d\varphi. \tag{23}$$

Here, $U(k_\rho,\varphi)$ can be thought of as a radial slice of $U(k_x,k_y)$ taken at an angle φ from the k_x-axis. If we rearrange this slightly to read

$$p(\rho=0,z\geq 0)=\frac{\omega_o\rho_o}{(2\pi)^2}\int_{0}^{k_o}\left[\frac{1}{2\pi}\int_{0}^{2\pi}U(k_\rho,\varphi;z=0)d\varphi\right]\times\frac{e^{iz\sqrt{k_o^2-k_\rho^2}}}{\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho, \tag{24}$$

then we recognize the quantity in brackets to be just the mean value of $U(k_\rho,\varphi;z=0)$ over its φ-dependence, which we will henceforth denote as $\langle U(k_\rho;z=0)\rangle_\varphi$. This allows us to write Eq. (24) as

$$p(\rho=0,z\geq 0)=\frac{\omega_o\rho_o}{(2\pi)^2}\int_{0}^{k_o}\langle U(k_\rho;z=0)\rangle_\varphi\frac{e^{iz\sqrt{k_o^2-k_\rho^2}}}{\sqrt{k_o^2-k_\rho^2}}k_\rho dk_\rho. \tag{25}$$

We are now able to use the constraint $k_o^2=k_\rho^2+k_z^2$ to change variables as before and write the axial pressure as

$$p(\rho=0,z\geq 0)=\frac{\omega_o\rho_o}{2\pi}\int_{0}^{k_o}\langle U(\sqrt{k_o^2-k_z^2};z=0)\rangle_\varphi e^{ik_z z}dk_z. \tag{26}$$

Equation (26) thus reveals how the propagating axial pressure arising from a two-dimensional velocity distribution *of arbitrary shape* can be directly computed via a one-dimensional inverse Fourier transform of the velocity spectrum on the aperture boundary averaged over its φ-dependence around the normal axis of observation.

As an informative example, we consider the case of a flat, rectangular piston undergoing unit-amplitude, harmonic motion in an infinite, rigid baffle, for which the velocity spectrum on the boundary takes the form[26]

$$U(k_x,k_y)=4\frac{\sin(k_x a)}{k_x}\frac{\sin(k_y b)}{k_y}. \tag{27}$$

If φ is taken to be the angle that k_ρ makes with the k_x-axis in the k_x-k_y plane, then this can be written in polar coordinates as

$$U(k_\rho,\varphi)\frac{4}{k_\rho^2}\frac{\sin(k_\rho a\cos\varphi)}{\cos\varphi}\frac{\sin(k_\rho b\sin\varphi)}{\sin\varphi}. \tag{28}$$

If (for simplicity), we further let $a=b=1$, and generate an array of values with $k_o=25$ for $U(k_\rho,\varphi)$, we can then easily compute the average value of the array over the φ-dimension. A discrete, inverse Fourier transform of the resulting one-dimensional array is then computed to produce the axial pressure profile shown in Fig. 11. Not surprisingly, the behavior of the pressure amplitude is similar to that of a baffled circular piston of comparable size (see Fig. 7), but demonstrates less pronounced nulls and virtually no Gibbs

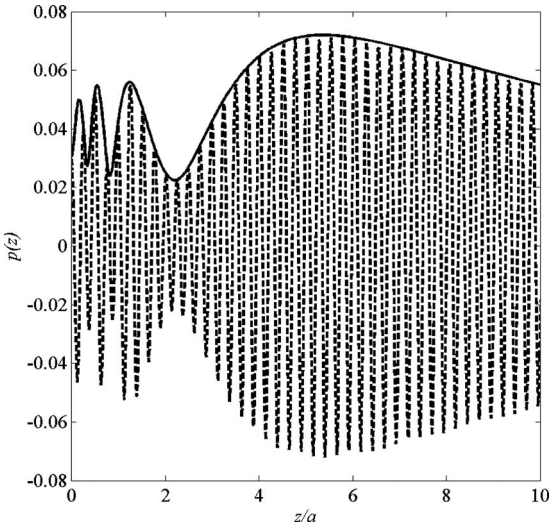


FIG. 11. Real part (dashed) and absolute value (solid) of complex axial pressure radiating from a baffled square piston for $k_0a=25$.

rippling, despite not having used a shading function in the computation (this would not be the case, in general; in this example, the axial spectrum simply happens to be very small near $k_z=0$). Unfortunately, unlike the circular piston, a full solution for the axial pressure associated with a rectangular piston is not available for comparison. However, as mentioned earlier, the technique introduced here does facilitate the derivation of an analytical solution for the propagating portion of the axial pressure.

We can obtain this solution by first noting that [in Eq. (27)] $\sin x$ can be written as

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{m=0}^{\infty} (-1)^m \frac{x^{2m+1}}{(2m+1)!}, \tag{29}$$

from which it follows that $\sin(\alpha x)/x$ takes the form

$$\frac{\sin(\alpha x)}{x} = \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m+1} x^{2m}}{(2m+1)!}. \tag{30}$$

With this result, the velocity spectrum can be expressed as

$$\begin{aligned} U(k_\rho,\varphi) &= \frac{4}{k_\rho^2} \sum_{m=0}^{\infty} (-1)^m \frac{(k_\rho a)^{2m+1} \cos(\varphi)^{2m}}{(2m+1)!} \\ &\quad \times \sum_{n=0}^{\infty} (-1)^n \frac{(k_\rho b)^{2n+1} \sin(\varphi)^{2n}}{(2n+1)!} \\ &= 4 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \frac{a^{2m+1} b^{2n+1} k_\rho^{2(m+n)}}{(2m+1)!(2n+1)!} \\ &\quad \times \cos(\varphi)^{2m} \sin(\varphi)^{2n}. \end{aligned} \tag{31}$$

The average value of this quantity with respect to the variable φ can now be evaluated by integrating from 0 to 2π, and making use of the integral formula

$$\int_0^{2\pi} \cos(\varphi)^{2m} \sin(\varphi)^{2n} d\varphi = 2B\left(m+\frac{1}{2}, n+\frac{1}{2}\right), \tag{32}$$

where $B(\alpha,\beta)$ is the Euler beta function.

The result is

$$\begin{aligned} \langle U(k_\rho; z=0)\rangle_\varphi &= 8 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \frac{a^{2m+1} b^{2n+1} k_\rho^{2(m+n)}}{(2m+1)!(2n+1)!} \\ &\quad \times B\left(m+\frac{1}{2}, n+\frac{1}{2}\right). \end{aligned} \tag{33}$$

Similar to Eq. (14) then, the axial pressure spectrum can be written as

$$\begin{aligned} P(k_z; \rho=0) &= \omega_o \rho_o \langle U(\sqrt{k_0^2 - k_z^2}; z=0)\rangle_\varphi \\ &= 8\omega_o \rho_o \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \\ &\quad \times \frac{a^{2m+1} b^{2n+1} (k_0^2 - k_z^2)^{(m+n)}}{(2m+1)!(2n+1)!} \\ &\quad \times B\left(m+\frac{1}{2}, n+\frac{1}{2}\right). \end{aligned} \tag{34}$$

Therefore, by Eq. (26), the propagating part of the axial pressure can be computed by evaluating the band-limited inverse Fourier transform of this equation

$$\begin{aligned} p(z; \rho=0) &= \frac{4\omega_o \rho_o}{\pi} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{m+n} \frac{a^{2m+1} b^{2n+1}}{(2m+1)!(2n+1)!} \\ &\quad \times B\left(m+\frac{1}{2}, n+\frac{1}{2}\right) \int_0^{k_0} (k_0^2 - k_z^2)^{m+n} e^{ik_z z} dk_z. \end{aligned} \tag{35}$$

The integral in this equation can be evaluated by making the substitution $k_z \to k_0 \cos\theta$, from which we obtain

$$\begin{aligned} &\int_0^{k_0} (k_0^2 - k_z^2)^{m+n} e^{ik_z z} dk_z \\ &= \int_0^{\pi/2} k_0^{2m+2n+1} \sin^{2m+2n} \theta e^{ik_0 z \cos\theta} \sin\theta d\theta \\ &= k_0^{2m+2n+1} \left[\int_0^{\pi/2} \cos(k_0 z \cos\theta) \sin^{2m+2n+1}\theta d\theta \right. \\ &\quad \left. + i \int_0^{\pi/2} \sin(k_0 z \cos\theta) \sin^{2m+2n+1}\theta d\theta \right]. \end{aligned} \tag{36}$$

One may recognize the last two terms as the integral representations of the Bessel and Struve functions, respectively.[27] Thus, the integral can be written as

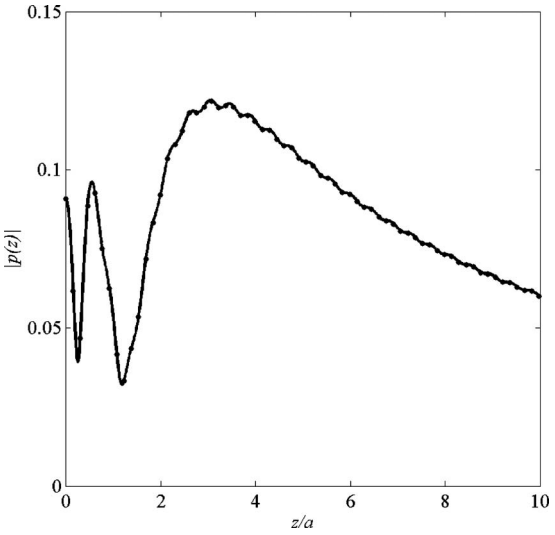


FIG. 12. Magnitude of the acoustic pressure propagating along the axis of a baffled square piston, obtained from a discrete implementation of Eq. (26) (dots) with $k_0a=15$ vs the theoretical result given by Eq. (38) (solid).

$$\int_0^{k_0} (k_0^2 - k_z^2)^{m+n} e^{ik_z z} dk_z = \frac{k_0^{2m+2n+1}\sqrt{\pi}\Gamma(m+n+1)}{2\left(\frac{z}{2}\right)^{m+n+1/2}} \times [J_{m+n+1/2}(k_0 z) + iH_{m+n+1/2}(k_0 z)] \tag{37}$$

[where $\Gamma(\alpha)$ is the Euler gamma function, $J_\nu(\alpha)$ is the Bessel function, and $H_\nu(\alpha)$ is the Struve function] so that the propagating part of the axial pressure can be written as

$$p(z;\rho=0) = \frac{2\omega_o\rho_o}{\sqrt{\pi}} \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} (-1)^{m+n} \left(\frac{2k_0^2}{z}\right)^{m+n+1/2} \times \frac{a^{2m+1}b^{2n+1}}{(2m+1)!(2n+1)!} B\left(m+\frac{1}{2}, n+\frac{1}{2}\right) \times \Gamma(m+n+1)[J_{m+n+1/2}(k_0 z) + iH_{m+n+1/2}(k_0 z)]. \tag{38}$$

This result represents an exact solution to Eq. (26), when $\langle U(\sqrt{k_o^2 - k_z^2}; z=0)\rangle_\varphi$ is given by Eq. (33). A computational evaluation of this formula can be problematic though, principally due to the factorial terms in the denominator, which can rapidly lead to a loss of accuracy as m and n become large. It was evaluated, however, for $k_0a=15$, the result of which is shown in Fig. 12, along with result obtained via the discrete Fourier transform approach. As expected, the two approaches yield the same result.

V. CONCLUSIONS

By way of the angular spectrum of plane waves representation of the pressure field diffracted by an axisymmetric two-dimensional aperture, we developed a direct mathematical relationship between the normal component of the acoustic velocity on the aperture boundary and the propagating portion of the acoustic pressure along the axis of symmetry. While, in its simplest form, this relationship is expressed in wavenumber space, we have seen that through the use of Fourier and Hankel transforms, one can obtain spatial representations of either of these quantities via numerical methods. We have further found that in implementing these methods, some care must be exercised in order to avoid the introduction of numerical errors associated with the conversion from a continuous to a discrete representation.

In terms of the axial pressure, this involved the application of a shading function to the argument of the discrete Fourier transform to achieve a smoothly terminating spectrum, and hence, to avoid the appearance of Gibbs phenomenon in the computed pressure values. Though not optimized, the approach was shown to enable us to characterize the axial pressure behavior of the aperture for values of z/a down to approximately 0.5; roughly an order of magnitude closer than an asymptotic approximation would permit. For values below this, the absence of higher frequency information prevents us from accurately characterizing the field. This approach was shown to hold for any two-dimensional aperture for which either the velocity boundary condition or its Hankel transform can be specified.

Alternatively, using the axial pressure to compute either the boundary condition or the far field diffraction pattern of the aperture was found to be more restrictive. As shown in the Appendix, these computations are only possible for problems in which the boundary condition on the plane of the aperture is axisymmetric, and known to be expressible as an entirely real quantity. While the example of the baffled piston is one such case, there are others of potential practical interest, such as certain axisymmetric vibrational modes of circular plates and membranes.

The efficiency afforded by this approach, however, may be its most redeeming feature. For instance, if one were to employ the conventional angular spectrum method to obtain the axial pressure associated with a rotationally symmetric aperture, the spectrum obtained by first computing the two-dimensional Fourier transform of the boundary condition would have to be propagated successively along the axis of symmetry and inverse transformed at each point to obtain one discrete value of the axial pressure. This new method, on the other hand, requires only a single Fourier transform to obtain the same set of axial pressure values after first computing the Hankel transform of the boundary condition. The computational savings, in this case, are evident since multiple two-dimensional Fourier transforms are reduced to a single one-dimensional Fourier transform.

ACKNOWLEDGMENTS

The author would like to thank Peter Stepanishen of the University of Rhode Island for his invaluable technical feedback in the preparation of this paper. It would certainly be of lesser quality without the benefit of his extensive experience in this field. This work was supported by the Naval Undersea Warfare Center, Newport, RI.

APPENDIX

From the form of Eq. (13) it is apparent that an evaluation of the Fourier transform on the right side of the equation will yield complex pressure values for all z. And though a relationship can be shown to exist between those pressure values corresponding to positive z, and those corresponding to negative z, the latter cannot be assumed to have any physical significance since by our definition of the boundary value problem, we have no knowledge of the conditions corresponding to negative z.

Thus, if we wish to compute the pressure spectrum in Eq. (14) (and subsequently the diffraction pattern) from values of the axial pressure, then we require pressure values for negative z as well as positive z. To see how the former values can be obtained, consider the simple example of a circular piston transducer vibrating at a fixed frequency, surrounded by an annular ring transducer vibrating with the same frequency, but slightly out of phase with the piston. In this type of situation, the specified boundary condition will be a complex quantity expressible as

$$\tilde{u}(\rho,z=0)=\mathrm{Re}[\tilde{u}(\rho,z=0)]+i\ \mathrm{Im}[\tilde{u}(\rho,z=0)], \tag{A1}$$

where the tilde is used here to explicitly denote complex quantities.

The Hankel transform of this quantity will also be complex, and can be written as

$$\tilde{U}(k_\rho,z=0)=\mathrm{Re}[\tilde{U}(k_\rho,z=0)]+i\ \mathrm{Im}[\tilde{U}(k_\rho,z=0)], \tag{A2}$$

where $\mathrm{Re}[\tilde{U}(k_\rho,z=0)]$ is the Hankel transform of $\mathrm{Re}[\tilde{u}(\rho,z=0)]$, and $\mathrm{Im}[\tilde{U}(k_\rho,z=0)]$ is the Hankel transform of $\mathrm{Im}[\tilde{u}(k_\rho,z=0)]$.

When this quantity is inserted into Eq. (13), we obtain

$$\tilde{p}(\rho=0,z)=\frac{\omega_o\rho_o}{2\pi}\int_{-\infty}^{\infty}\mathrm{Re}[\tilde{U}(k_\rho,z=0)]H(k_z)e^{ik_zz}dk_z + i\frac{\omega_o\rho_o}{2\pi}\int_{-\infty}^{\infty}\mathrm{Im}[\tilde{U}(k_\rho,z=0)]H(k_z)e^{ik_zz}dk_z. \tag{A3}$$

So, since each of these two integrals represents the inverse Fourier transform of an asymmetric, real quantity, the result of each will be a Hermitian quantity, which we write as

$$\begin{aligned}\tilde{p}(\rho=0,z)&\equiv\tilde{p}_1(z)+i\tilde{p}_2(z)\\&=\mathrm{Re}[\tilde{p}_1(z)]+i\ \mathrm{Im}[\tilde{p}_1(z)]+i\ \mathrm{Re}[\tilde{p}_2(z)]\\&\quad-\mathrm{Im}[\tilde{p}_2(z)],\end{aligned} \tag{A4}$$

where the first and third terms (i.e., the real terms) are even functions of z, and the imaginary terms are odd.

Equation (A4) reveals the form of the complex axial pressure produced by evaluating the inverse Fourier transform of Eq. (13) for a complex boundary condition $\tilde{u}(\rho,z=0)$. The term $\tilde{p}_1(z)$ is that portion of the axial pressure produced by the real part of the boundary condition, whereas $\tilde{p}_2(z)$ is the portion produced by the imaginary part of the boundary condition. Thus, if the boundary condition is a purely real quantity, then the resulting complex axial pressure simply becomes the sum of an even, real function and its Hilbert transform. Computationally (or experimentally), this is readily accommodated. Having computed (or measured) the real axial pressure values for positive z, one simply adds the Hilbert transform and folds the conjugate of the result over to the negative z-axis. This operation yields a set of complex axial pressure values for both positive and negative z, consistent with Eq. (13), and therefore permits the Fourier transform leading to Eq. (14).

If we now consider the general case of a complex boundary condition, then by Eq. (A4), the real part of the complex axial pressure predicted by Eq. (13) assumes the form

$$\mathrm{Re}[\tilde{p}(\rho=0,z)]=\mathrm{Re}[\tilde{p}_1(z)]-\mathrm{Im}[\tilde{p}_2(z)]. \tag{A5}$$

Here, the pressure is seen to be the difference between an even and odd function. Such a difference results in an overall function that, in general, has no symmetry with respect to z. As such, the real pressure values measured along the positive z-axis cannot be used to construct values corresponding to the negative z-axis. To reiterate then, the diffraction pattern of an axisymmetric aperture can only be obtained from values of the axial pressure if the aperture function is expressible as a purely real (or purely imaginary) quantity, in general.

[1] G. C. Sherman, "Application of the convolution theorem to Rayleigh's integral formulas," J. Opt. Soc. Am. **57**, 546–547 (1967).
[2] H. Weyl, "Ausbreitung electromagnetischer Wellen ber einem ebenen Leiter (Propagation of electromagnetic waves over a conducting plane)," Ann. Phys. **365**(21), 481–500 (1919).
[3] P. C. Clemmow, "A method for the exact solution of a class of two-dimensional diffraction problems," Proc. R. Soc. London, Ser. A **205**, 286–308 (1951).
[4] G. C. Gaunaurd and H. Uberall, "Acoustics of finite beams," J. Acoust. Soc. Am. **63**, 5–16 (1978).
[5] P. V. Chitnis, P. E. Barbone, and R. O. Cleveland, "Customization of the acoustic field produced by a piezoelectric array through interelement delays," J. Acoust. Soc. Am. **123**, 4174–4185 (2008).
[6] X. Zeng and R. J. McGough, "Evaluation of the angular spectrum approach for simulations of near-field pressures," J. Acoust. Soc. Am. **123**, 68–76 (2008).
[7] X. Zeng and R. J. McGough, "Optimal simulations of ultrasonic fields produced by large thermal therapy arrays using the angular spectrum approach," J. Acoust. Soc. Am. **125**, 2967–2977 (2009).
[8] G. Zhou, "Far-field structure of a linearly polarized plane wave diffracted by a rectangular aperture," Opt. Laser Technol. **41**, 504–508 (2009).
[9] J.-H. Li, Y.-W. Cheng, Y.-C. Chue, C.-H. Lin, and T. W.-H. Sheu, "The influence of propagating and evanescent waves on the focusing properties of zone structure plates," Opt. Express **17**, 18462–18468 (2009).
[10] K. Matsushima and T. Shimobaba, "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," Opt. Express **17**, 19662–19673 (2009).
[11] S. H. Kazemi, M. M. Mirsalehi, and A. R. Attari, "Comparison of iterative angular spectrum and optimal rotation angle methods in designing beam-fanners," Opt. Express **17**, 14825–14831 (2009).
[12] T. B. Hansen and A. D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields* (IEEE, New York, 1999).
[13] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. (Academic, San Diego, CA, 1995).
[14] A. F. Medeiros and P. R. Stepanishen, "The forward and backward projection of acoustic fields from axisymmetric ultrasonic radiators using impulse response and Hankel transform techniques," J. Acoust. Soc. Am. **75**, 1732–1740 (1984).
[15] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).

[16]L. V. King, "On the acoustic radiation field on the piezo-electric oscillator and the effect of viscosity on transmission," Can. J. Res. **11**, 135–155 (1934).
[17]M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, London, 1999).
[18]A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1987).
[19]P. R. Stepanishen, "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[20]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (The Acoustical Society of America, Melville, NY, 1989).
[21]D. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[22]J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1996).
[23]S. Pan and A. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," IEEE Trans. Acoust. Speech Signal Process. **31**, 1262–1275 (1983).
[24]J. Blackledge, *Quantitative Coherent Imaging Theory and Methods and Some Applications* (Academic, San Diego, CA, 1989).
[25]N. Bose, *Digital Filters Theory and Applications* (North-Holland, New York, 1985).
[26]D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[27]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1964).

Nondiffuse elastic and anelastic passive imaging

Francesco Mulargia[a] and Silvia Castellaro
Dipartimento di Fisica, Settore di Geofisica, Università di Bologna, Bologna 40121, Italy

(Received 24 July 2009; revised 23 December 2009; accepted 5 January 2010)

The property at the basis of passive acoustic imaging is that, taken any two points, one of them can be seen as the source of the waves and the other as the recording station. This property, which was shown to hold also in nondiffuse fields, is here exploited: (1) to allow an undistorted passive imaging through the simple use of the statistical mode to estimate wave velocity, (2) to determine the azimuth of the instantaneous Huygens sources of the noise wavefield, and (3) to measure, provided that the noise bandwidth is wide with respect to that of the local system, the material dissipation constant as a function of frequency. The authors applied this theory to study the seismic noise field in the Ravenna, North-Central Italy, shore area and found it capable to provide velocity dispersion curves matching those of independent surveys, to track the sources of seismic noise to a few major firms in Ravenna port, with the prevailing source switching at the time scale of seconds, and to measure the dissipation quality factor Q at ~20 independent of frequency in the range 1–30 Hz. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3298939]

PACS number(s): 43.60.Lq, 43.60.Gk, 43.60.Jn, 43.80.Lb [RKS] Pages: 1391–1396

I. INTRODUCTION

The analysis of the spatial correlation of seismic noise has been recently formalized within the general framework of the theory of acoustic diffuse fields[1–5] and shown capable to provide a detailed imaging of Earth's crust.[6] This is somehow surprising since the diffuse field approach requires strong constraints hardly realized in a seismic noise wavefield, such as modal energy equipartition,[7–9] which is only encountered in highly fractured and nondissipative subsoil environments.[10] Such a passive imaging capability of noise spatial correlation appears more simply to lie in the fact that cross-correlating the noise field at two different spatial points leads to a nonuniform mapping in apparent velocity. In particular, a field with wave velocity uniformly distributed with azimuth θ over the angular aperture $(\theta_2-\theta_1)$, but with no constraint on wave amplitude (cf. Fig. 1), is mapped into a field with probability density function g strongly nonuniform in apparent velocity[11]

$$g(\Xi)=\left|\frac{1}{(\theta_2-\theta_1)}\frac{1}{\Xi\sqrt{(\Xi^2-1)}}\right|, \tag{1}$$

where Ξ is the normalized apparent velocity, i.e., the inverse of the adimensional azimuthal slowness

$$\Xi=\frac{v_{\text{apparent}}}{v_{\text{true}}}=\frac{1}{\cos(\theta)}. \tag{2}$$

For any finite field aperture $0°<(\theta_2-\theta_1)\leq 360°$ and considering this field explored at the two points A and B, the distribution $g(\Xi)$ is strongly peaked at $\Xi=1$; i.e., the cross-correlation between the recorded waveforms at A and at B is dominated by waves traveling along the array AB in either direction. According to Huygens principle, A or B will act as point sources and, respectively, B or A as recording stations, allowing to derive the elastodynamic retarded and advanced Green's functions by definition.[11] In the case of an azimuthally isotropic wavefield the latter functions will be identical in concordance with the general property of diffuse fields,[11] granted by reciprocity.[12]

The present work aims at exploiting the above statistical-ballistic property and to show that this property allows the following.

(1) An unbiased measure of wave velocity (both phase velocity, by considering the onset, and group velocity, by introducing an appropriate filter to locate the centroid of the wavepacket[13]). The deployment of semicircular arrays together with the use of the statistical mode as an estimator provides, in fact, an undistorted passive imaging without the need of an explicit correction for astigmatism such as the one suggested by Mulargia and Castellaro.[11]
(2) The azimuthal identification of the instantaneous Huygens sources of the noise wavefield and, under favorable conditions, to ascertain the origin of each of them.
(3) A measure of the material dissipation quality factor Q as a function of frequency, provided that the noise bandwidth is wide compared to the response of the local system in the neighborhood of the points A and B.

II. THE STATISTICAL MODE AS UNBIASED Ξ ESTIMATOR

Consider a wavefield composed of plane waves with uniformly distributed random azimuth in a given angular interval $\theta_1\leq\theta\leq\theta_2$, while no constraint is imposed on the distribution of wave amplitude, which can be of any functional form (see Fig. 1). Let us measure the velocity of such waves by considering in a given time interval T the cross-correlation R_{OS_j} at the radial two-point couples of a semicircular array composed of $n+1$ stations with center in O and peripheral stations S_j (Fig. 2).

[a]Author to whom correspondence should be addressed. Electronic mail: francesco.mulargia@unibo.it

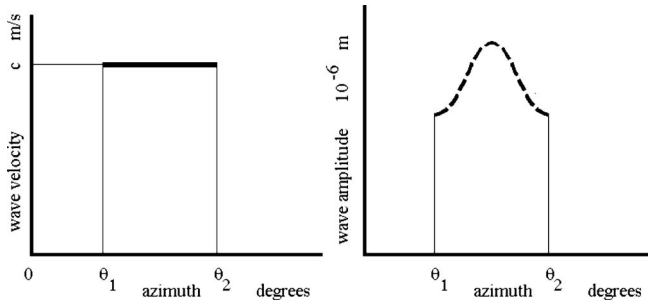


FIG. 1. The wavefields considered here have (left) a uniform probability density function $f(x)=1/(\theta_2-\theta_1)$ on velocity in the angular domains $\theta_1 \le \theta \le \theta_2$ and $0 \le \theta_1 \le \theta_2 \le \pi$. No constraint (right) is imposed on the distribution of wave amplitude, which is here pictorially shown as a Gaussian beam.

Using the above theory, the apparent normalized wave velocity Ξ on a number r of radial station couples will fall near the peak of $g(\Xi)$. The latter couples measure nearly identical Ξ values, thus giving the definition of the statistical mode as the most probable value of a set of random variables.[14]

Given the definition of Ξ [Eq. (2)] in a uniform wavefield, the values of Ξ determining the statistical mode will occur for the angularly consecutive radial station couples $O-S_k, O-S_{k+1}, O-S_{k+2}, \ldots, O-S_{k+r}$ (Fig. 2 middle) as follows:

$$|\Xi_{j=k} - \Xi_{j=k+d}| \simeq 0, \quad d = 0,1,2, \ldots, r. \tag{3}$$

We may therefore define the mode of the normalized apparent velocity as

$$\text{mode}(\Xi) = \Xi_{j=k+d}, \quad d = 0,1,2, \ldots, r. \tag{4}$$

Note that the use of such an array introduces a discrete sampling of the field, which results in a coarsening $\Delta\phi$ inversely proportional to station number. The maximum misalignment that each ray may have with the closest oriented radial couple $O-S_j$ is (see Fig. 2)

$$\Delta\phi = \frac{\pi}{2(n-1)}. \tag{5}$$

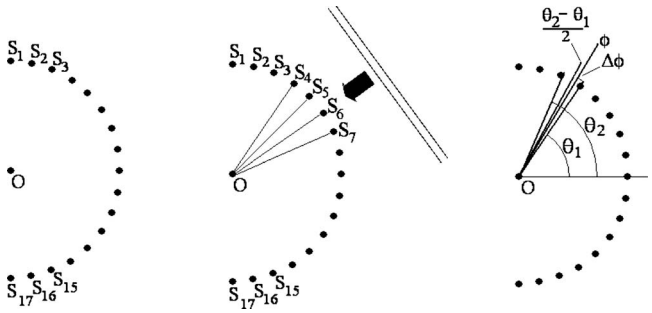


FIG. 2. The semicircular array with center in O, depicted in the case with $n=17+1$ stations S_j (left). The measurement of velocity by cross-correlation on radial station couples $O-S_j$ yields that the statistical mode is defined by the radial array couples most closely aligned with the field. The case of a plane wave is shown, together with the most closely aligned radial array couples $O-S_4/O-S_7$ (middle). Each array implies an azimuthal discretization coarsening with a maximum misalignment $\Delta\phi$ equal to $|\theta_2-\theta_1|/2$ (right).

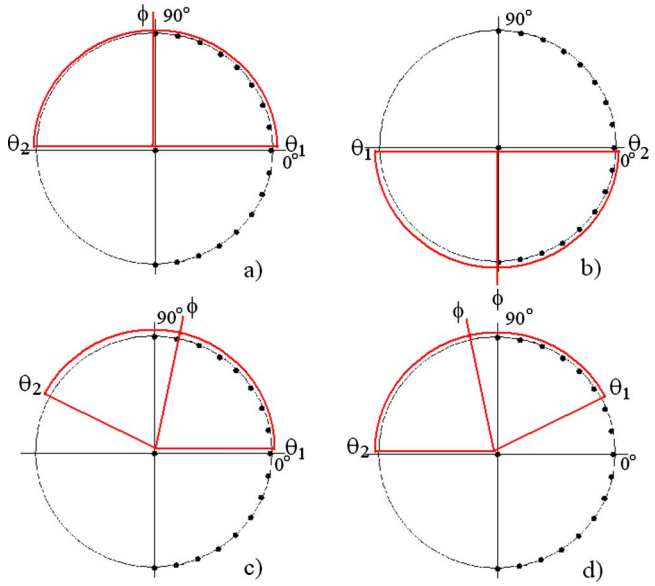


FIG. 3. (Color online) The field geometry: the limiting azimuths θ_1 and θ_2 and the central azimuth ϕ. The cases shown are (a) uniform{[0°;180°]}, $\phi=90°$; (b) uniform{[180°;360°]}, $\phi=270°$; (c) uniform{[0°;158°]}, $\phi=79°$; and (d) uniform{[22°;180°]}, $\phi=101°$. The angles are measured according to the standard trigonometric notation.

Let us study how the statistical mode depends on field aperture $\theta_1 \le \theta \le \theta_2$ and azimuth ϕ by analyzing a large number N of realizations of uniform wavefields in which the number of stations n of a semicircular radial array is kept constant. In this study we use $N > 10^7$, $n=17+1$, and explore a variety of cases of $\theta_1 \le \theta \le \theta_2$ and ϕ.

Consider first of all the case of an azimuthally isotropic field, θ:uniform{[0;360°]}. A bimodal Ξ distribution [Fig. 4(a)] is found, with mode Ξ values equal to -1 and 1. As recalled above, the latter represent two equivalent solutions, corresponding by definition, respectively, to the advanced and retarded Green functions.[11] Note how this reflects the symmetry with respect to time that characterizes diffuse fields, and note also that the "forbidden" region between -1 and 1 [deriving from $\cos(\theta) \le 1$ in Eq. (2)] stands for a bias on Ξ, which is always positive; i.e., the value of the apparent velocity is never lower than the true value in either direction of propagation along the array AB.

Let us next consider the simulation of azimuthally anisotropic wavefields, and start by observing that both the half-space fields uniform{[0;180°]} and uniform{[180;360°]}, i.e., respectively, with orthogonal azimuth 90° and 270° [see Figs. 3(a) and 3(b)] produce a pattern identical to that of the azimuthally isotropic case [Fig. 4(a)]. This is due to the numerical balance between the rays crossing radially the array in direction S_j-O and those crossing it in direction $O-S_j$. Such a balance occurs for arrays with azimuth $\phi=90°$ or $\phi=270°$ irrespective of aperture, and a bimodal symmetric distribution peaked at 1 and -1 is found in each case. The independence of wavefield aperture is illustrated by the extreme case of a zero aperture field uniform{[90°;90°]}, i.e., a plane wave with azimuth $\phi=90°$, which still produces a pattern identical to that of the azimuthally isotropic case [Fig. 4(a)].

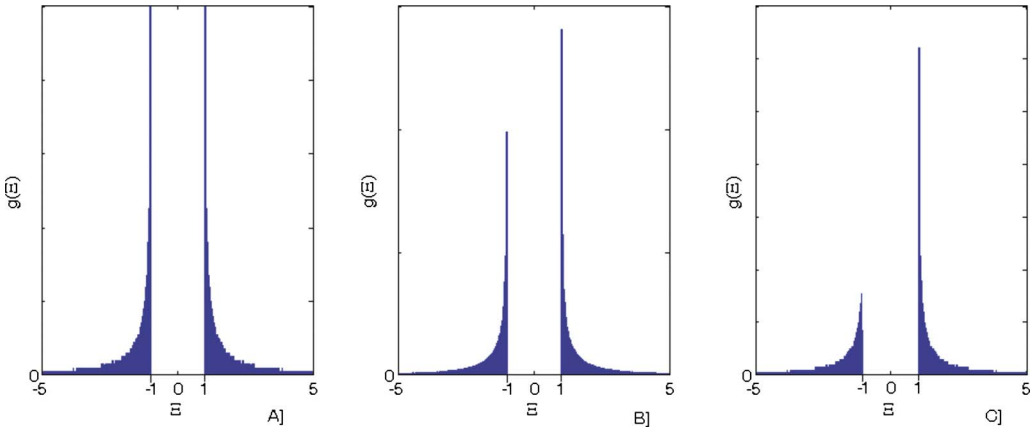


FIG. 4. (Color online) The probability density function of the normalized apparent velocity for [*A*] an azimuthally isotropic field or for the orthogonal fields uniform{[0;180°]} and uniform{[180;360°]}, [*B*] θ:uniform{[0;158°]}, and [*C*] θ:uniform{[0;152°]}. Fields with the same azimuth but smaller apertures lead the lower amplitude statistical mode to visually disappear and are not shown. Two ticks at $|\Xi|=1$, corresponding to true velocity, are also shown.

The inward/outward radial balance is rapidly lost if the field azimuth is rotated from 90° or 270°. For example, the uniform fields with aperture 158° and azimuth $\phi=76°$ [see Fig. 3(c)] result still in a bimodal $g(\Xi)$ distribution [Figs. 4(b) and 4(c)], but with one prevailing mode. The wave is seen as still traveling in both directions, and both the advanced and retarded Green functions can be recovered, but one has a higher intensity and the picture is no more time invariant.[11] Note that fields with the same azimuth but smaller apertures lead the lower amplitude statistical mode to visually disappear and are not shown.

In other words, Ξ is peaked at either -1 or at 1 depending on the fact that "more rays" cross the array radially inward or outward. The variation of ϕ above and below 90° and 270° rules this inward/outward balance, and, if aperture remains the same, it has the only effect of switching the prevailing mode. For example, the field uniform{[0°;158°]}, $\phi=79°$ and its specular θ:uniform{[−22°;180°]}, $\phi=101°$ [Figs. 3(c) and 3(d)] have specular bimodal density functions.

We may therefore conclude that, independent of wavefield aperture and azimuth and even in the extreme case of a single plane wave, the estimate of scalar wave velocity using the statistical mode over all radial couples is always unbiased.

It must be noted that the coarsening introduced by station number results in a lower accuracy in the measure of Ξ, which reflects in the mode. From Eq. (5) this inaccuracy is equal to

$$\Delta\Xi = \frac{1}{\cos(\pi/[2(n-1)])}, \tag{6}$$

so that an array with $n=17+1$ stations will have a (positive) coarsening bias $\Delta\Xi<2\%$ and an array with $n=7+1$ stations will have a (positive) coarsening bias $\Delta\Xi<4\%$.

It must also be noted that using the mean of Ξ as a location parameter instead of the mode would always give biased estimates. In fact, the mean is the first moment of the Ξ distribution[11]

$$E(\Xi) = \frac{\log(\Xi+\sqrt{\Xi^2-1})}{\sec^{-1}(\Xi_2)-\sec^{-1}(\Xi_1)} \quad I_{\Xi}[\sec^{-1}(\Xi_1),\sec^{-1}(\Xi_2)], \tag{7}$$

which is clearly unbounded at $k\pi/2$, with $k=1,2,\ldots$. The mean of Ξ in azimuthal domains not comprising $k\pi/2$ is bounded but biased, and with a bias always larger than that of the mode. The use of the median as a location parameter, placing the central value in the middle of the distribution domain, would also provide biased estimates, albeit with a generally smaller bias than the mean.[15] Note how in all cases the bias would be positive.

The above argument can be summarized as follows. In a radial semicircular array the statistical mode of the normalized apparent velocity Ξ measured over the radial station couples provides always an unbiased estimate of wave velocity on wavefields with azimuthally uniform velocity distribution and whatever aperture and intensity. The efficiency of the statistical mode as an estimator provides the weight factor that guarantees the effectiveness of Aki's spatial autocorrelation[16] technique. At the same time, the peak of $g(\Xi)$ is always placed in correspondence of the minimum norm of Ξ, so that the mode of Ξ coincides with the minimum, an issue somehow empirically acknowledged by the refraction microtremor technique.[17]

III. AZIMUTHAL IDENTIFICATION OF THE INSTANTANEOUS HUYGENS SOURCES

In light of the above discussion, the Ξ distribution is always bimodal asymmetric, with symmetry achieved only by isotropic fields and orthogonal fields. Hence, the probability density function $g(\Xi)$ will generally provide one prevailing mode, which will reflect the dominant instantaneous Huygens source(s). As a consequence, splitting a recorded signal of length T into $N=T/\Delta t$ subsequent subintervals, each of these will be characterized by a probability density function with a prevailing mode and the azimuth of the instantaneous Huygens source in each subinterval will be sim-

ply given by the central azimuth ϕ_{mode} of the sector corresponding to the mode Ξ (cf. Fig. 2 middle) as follows:

$$\phi_{\text{mode}} = \frac{\phi_k + \phi_{k+r}}{2}. \tag{8}$$

IV. ESTIMATE OF THE MATERIAL DISSIPATION CONSTANT

A well known result in the theory of dynamical systems is that the response $y(t)$ of a linear time invariant system can be written as the convolution

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)du, \tag{9}$$

where $x(t)$ is the input signal and h is the system response to a unit impulse (i.e., Green's function). By definition of the input-output cross-correlation $R_{xy}(\tau)$

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t-\tau)y(t)dt \tag{10}$$

and therefore

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t-\tau)\int_{-\infty}^{\infty} h(u)x(t-u)dudt. \tag{11}$$

Hence, interchanging the order of integration

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(u)\left[\lim_{T\to\infty} \frac{1}{T}\int_0^T x(t-\tau)x(t-u)dt\right]du, \tag{12}$$

and, after a change in variables,

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} h(u)R_{xx}(\tau-u)du, \tag{13}$$

where R_{xx} is the autocorrelation of the input.

If the noise bandwidth is large with respect to the impulse response of the local mechanical system, i.e., the material between the two points A and B considered, h can be taken as constant and factored so that, for values of t around τ,

$$R_{xy}(\tau) = h(\tau)\int_{\infty}^{\infty} R_{xx}(\tau-u)du, \tag{14}$$

yielding that the cross-correlation of the displacement at A and B will be directly proportional to the impulse system response. This result, ascribed to equipartition by the diffuse field theory, thanks to the above statistical-ballistic property applies for example to seismic noise, which has a very large bandwidth extending from 10^{-5} to 10^{3} Hz.[18]

This allows estimating not only the elastic but also the anelastic material properties from cross-correlation recorded over a commendably long period of time to guarantee wide bandwidth. In fact, the quality factor Q, the reciprocal of the specific dissipation constant, is defined as a function of the relative energy loss per cycle at the frequency ν normalized to the absolute energy $\Delta E(\nu)/E(\nu)$ as

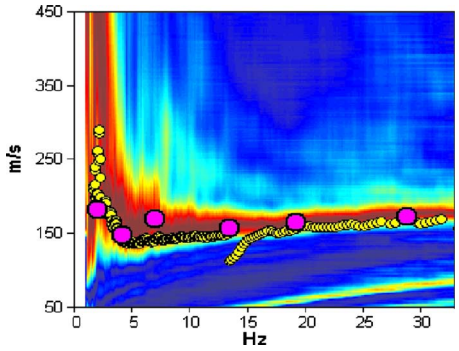


FIG. 5. (Color online) The local elastic wave velocity dispersion at the Marina di Ravenna experimental site measured by using the independent estimates of ReMi (contour plot) and ESAC (yellow circles — light gray in b/w) and by using the azimuthal mode (magenta circles — dark gray in b/w) as described by the present work.

$$\frac{\Delta E(\nu)}{E(\nu)} = \frac{2\pi}{Q(\nu)}, \tag{15}$$

which is the tangent of the phase angle by which the strain lags behind the applied stress in the case of sinusoidal excitation. In light of the above discussion, Q can be measured as a function of frequency ν both in the frequency domain using cross-spectral density and, after bandpass filtering, directly in the time domain from the cross-correlation function.

We tested the above theory through an experiment on seismic noise.

V. EXPERIMENTAL VALIDATION

Our experiment was designed to minimize two effects, which are spurious to the present problem. First, we aimed at minimizing the effect of lateral heterogeneities, which may act as multiple scatterers of unknown position making it impossible to check the capability of the above formulation to effectively locate the Huygens sources. In a homogeneous subsoil, the Huygens sources are true sources, i.e., the points where the beams originate before hitting the stations, which makes this check possible. Second, we aimed at minimizing the effect of subsoil layering, which may induce the excitation of higher order wave eigenmodes, thus introducing further an unnecessary variability.

To this extent, we selected a site in North-Central Italy, on the Marina di Ravenna beach at approximately 150 m from the Adriatic Sea shore, where the superficial layer of sand is virtually isotropic and homogeneous down to a depth of the order of 10^2 m and for a horizontal extension of several tens of kilometers.[19]

The validity of the subsoil homogeneity and isotropy approximations in terms of seismic wave propagation was first of all verified through a refraction microtremor[17] and an extended spatial autocorrelation[20] analysis by using a Micromed Rosina 24 bit resolution digital array with 16 vertical 4.5 Hz geophones arranged, respectively, with a 12+4 L-shaped geometry, a 5 m interdistance, and using a 1024 Hz sampling rate. The two analyses showed a remarkably narrow scatter and a virtual absence of dispersion (Fig. 5) in the frequency range 1–40 Hz, with a nearly constant phase ve-

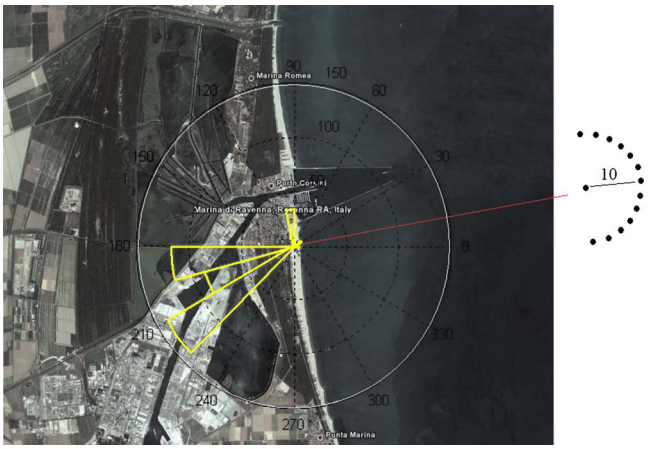


FIG. 6. (Color online) The azimuthal distribution of noise over the whole measurement period of 820 s as measured with the semicircular array with 10 m radius oriented as shown at the right of the figure and positioned at the center of the rose. The subsoil homogeneity makes these real sources rather than scatterers and allows identification as sources some specific industrial firms of the Ravenna port.

locity of 160 m/s. This stands for a half-space configuration well approximated over wavelengths $\lambda \simeq 4-160$ m.

We then started the test of the above theory by deploying the digital array Rosina in a semicircular layout using 13+1 vertical sensors with a 10 m radius (Fig. 6). A recording of seismic noise 820 s long was taken on calm sea conditions. The dispersion of phase velocity was measured in the time domain through a correlation analysis following 1/7 octave 40 dB narrow band filtering centered at the frequencies of 1, 3, 7, 13, 19, and 29 Hz. The cross-correlation was then performed on all the array radial couples, and the scalar velocity estimated from the mode Ξ over the whole 820 s recording time interval. The results are reported by the pink dots in Fig. 5, and were found to be in good agreement with the independent extended spatial autocorrelation and refraction microtremor measurements performed in our preliminary analysis.

The existence of horizontally traveling noise waves with a vertical displacement component along the surface of a nearly isotropic and homogeneous material suggests that they are Rayleigh waves in the fundamental eigenmode.[21]

We then proceeded to determine the instantaneous Huygens sources. To this extent, the recording was split into 820 nonoverlapping subintervals 1 s long, and central azimuth corresponding to the mode Ξ in each subinterval was determined according to Eq. (8). This analysis showed that seismic noise originated from three main azimuths at 190°, 220°, and 100° (Fig. 6). These, in light of the lack of subsoil scatterers, should be identified as three true sources, a result which appeared consistent with reality, since they were pointing to the industrial area of the Ravenna port.

We then plotted the mode azimuth versus time to characterize the dynamics of the seismic noise field, and found (Fig. 7) that the source azimuth varied randomly and abruptly at the time scale of a few seconds—the persistence time varied from 1 to 10 s—suggesting that instantaneous Huygens source interference[5] of impulsive wavepackets is at the origin of seismic noise.

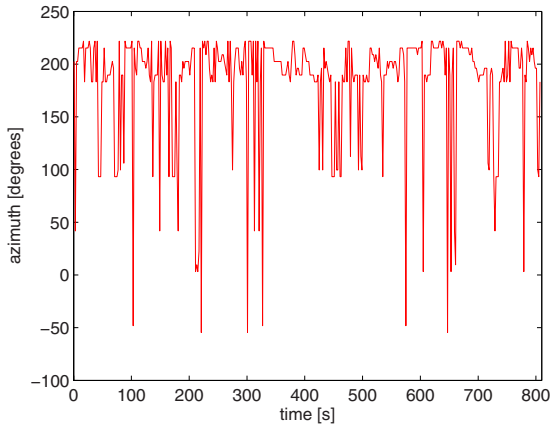


FIG. 7. (Color online) The azimuthal distribution of noise versus time during the measurement period of 820 s.

Finally, we estimated the anelastic constant Q according to Eq. (15). For compatibility with the previous dispersion analysis, the estimate was made in the time domain, using the whole T recording interval together with the 1/7 octave 40 dB narrow band filtering centered at the frequencies of 1, 3, 7, 13, 19, and 29 Hz. This resulted to nearly frequency independent of $Q \simeq 15-20$ (see Fig. 8). In light of the Rayleigh nature of the noise wavepackets, this provides an estimate of the shear wave Q_S.[22]

VI. DISCUSSION AND CONCLUSIONS

Extending our previous work on passive acoustic imaging on nondiffuse fields, we show here that undistorted elastic imaging is possible through the simple use of the statistical mode, instead of the mean or the median, to estimate wave velocity from noise correlation at different radial couples of a semicircular array. This provides an immediate measure, which in turn allows azimuthally locating the instantaneous Huygens sources of noise. At the same time, since seismic noise bandwidth is generally much wider than the response of the medium between the correlated points, a passive measure of the anelastic dissipation constant is also

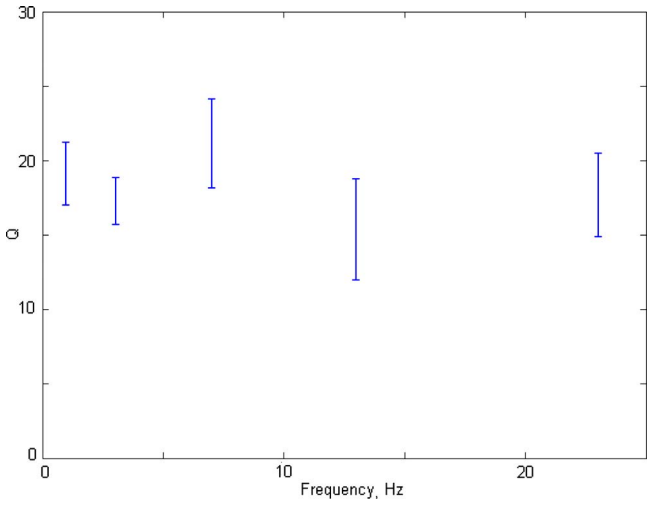


FIG. 8. (Color online) The local anelastic parameter Q as a function of frequency.

possible. The theoretical predictions were confirmed by an experiment performed on a Ravenna beach in North-Central Italy, which showed how this approach yields measured values of the dispersion of the phase velocity of Rayleigh waves identical to those obtained with established techniques, allowing the simultaneous identification of the instantaneous sources of noise and their time evolution. In such experiment, the sources of seismic noise were tracked to three major firms in the Ravenna port, and the prevailing source was found to switch at the time scale of seconds, suggesting instantaneous impulsive wavepacket interference as the basic mechanism of seismic noise. A passive measure of the anelastic dissipation was also obtained, with values of the quality factor $Q \simeq 15-20$ approximately independent of frequency in the range 1–30 Hz, but quite unfortunately no independent measurements could be found for this parameter, making a cross-check of the measured values impossible.

ACKNOWLEDGMENT

This work was performed with PRIN PTRC4C-004 grant.

[1]R. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[2]O. Lobkis and R. Weaver, "On the emergence of Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[3]R. Weaver and O. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[4]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[5]R. Snieder, J. Shieman, and R. Calvert, "Equivalence of the virtual source method and wave-field deconvolution in seismic interferometry," Phys. Rev. E **73**, 066620 (2006).
[6]M. Campillo and A. Paul, "Long range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[7]R. Sax and R. Hartenberger, "Theoretical prediction of seismic noise in a deep borehole," Geophysics **29**, 714–720 (1964).
[8]T. Tanimoto, "The three-dimensional shear wave structure in the mantle by overtone waveform inversion—I. Radial seismogram inversion," Geophys. J. R. Astron. Soc. **89**, 713–740 (1987).
[9]B. Romanowicz, *International Handbook of Earthquake and Engineering Seismology* (Academic, New York, 2003), Chap. II.
[10]R. Hennino, N. Tregoures, N. Shapiro, L. Margerin, M. Campillo, B. van Tiggelen, and R. Weaver, "Observation of equipartition of seismic waves," Phys. Rev. Lett. **86**, 3447–3450 (2001).
[11]F. Mulargia and S. Castellaro, "Passive imaging in nondiffuse acoustic wavefields," Phys. Rev. Lett. **100**, 218501 (2008).
[12]A. Mandelis, *Diffusion-Wave Fields: Mathematical Methods and Green Functions* (Springer, New York, 2001).
[13]A. L. Levshin, V. F. Pisarenko, and G. A. Pogrebinsky, "On a frequency-time analysis of oscillations," Ann. Geophys. **28**, 211–218 (1972).
[14]E. Parzen, "On estimation of a probability density function and mode," Ann. Math. Stat. **33**, 1065–1076 (1962).
[15]A. Mood, F. Graybill, and D. Boes, *Introduction to the Theory of Statistics*, 3rd ed. (McGraw-Hill, New York, 1974).
[16]K. Aki, "Space and time spectra of stationary stochastic waves with special reference to microtremors," Bull. Earthquake Res. Inst., Univ. Tokyo **35**, 415–457 (1957).
[17]J. Louie, "Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor arrays," Bull. Seismol. Soc. Am. **91**, 347–364 (2001).
[18]R. Widmer-Schnidrig, "What can superconducting gravimeters contribute to normal-mode seismology?," Bull. Seismol. Soc. Am. **93**, 1370–1380 (2003).
[19]A. Amorosi, M. Centineo, M. Colalongo, and F. Fiorini, "Millennial-scale depositional cycles from the Holocene of the Po Plain, Italy," Mar. Geol. **7**, 222–223 (2005).
[20]M. Ohori, A. Nobata, and K. Wakamatsu, "A comparison of ESAC and f-k methods of estimating phase velocity using arbitrarily shaped microtremor arrays," Bull. Seismol. Soc. Am. **92**, 2323–2332 (2002).
[21]K. Aki and P. Richards, *Quantitative Seismology*, 2nd ed. (University Science Books, Sausalito, CA, 2002).
[22]A. Ben-Menahem and S. Singh, *Seismic Waves and Sources*, 2nd ed. (Dover, New York, 1998).

Middle ear function and cochlear input impedance in chinchilla

Michaël C. C. Slama
Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Michael E. Ravicz and John J. Rosowski
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114

(Received 31 July 2009; revised 2 December 2009; accepted 7 December 2009)

Simultaneous measurements of middle ear-conducted sound pressure in the cochlear vestibule $\mathbf{P_V}$ and stapes velocity $\mathbf{V_S}$ have been performed in only a few individuals from a few mammalian species. In this paper, simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ in six chinchillas are reported, enabling computation of the middle ear pressure gain $\mathbf{G_{ME}}$ (ratio of $\mathbf{P_V}$ to the sound pressure in the ear canal $\mathbf{P_{TM}}$), the stapes velocity transfer function **SVTF** (ratio of the product of $\mathbf{V_S}$ and area of the stapes footplate A_{FP} to $\mathbf{P_{TM}}$), and, for the first time, the cochlear input impedance $\mathbf{Z_C}$ (ratio of $\mathbf{P_V}$ to the product of $\mathbf{V_S}$ and A_{FP}) in individuals. $|\mathbf{G_{ME}}|$ ranged from 25 to 35 dB over 125 Hz–8 kHz; the average group delay between 200 Hz and 10 kHz was about 52 μs. **SVTF** was comparable to that of previous studies. $\mathbf{Z_C}$ was resistive from the lowest frequencies up to at least 10 kHz, with a magnitude on the order of 10^{11} acoustic ohms. $\mathbf{P_V}$, $\mathbf{V_S}$, and the acoustic power entering the cochlea were good predictors of the shape of the audiogram at frequencies between 125 Hz and 2 kHz.
 [DOI: 10.1121/1.3279830]

PACS number(s): 43.64.Ha, 43.80.Lb [BLM] Pages: 1397–1410

I. INTRODUCTION

In terrestrial vertebrates, sound pressure waves are transmitted from the outside world to the fluid-filled inner ear primarily via the tympanic membrane (TM) and the ossicular chain of the middle ear. In mammals, the ossicular chain is made of three small bones: the malleus, the incus, and the stapes. Motion of the stapes footplate in the oval window produces pressure waves in the perilymph of the vestibule, the entrance to the cochlea. These waves propagate in the cochlea and are sensed by the organ of Corti, transduced into neural impulses, and sent to the central auditory system via the auditory nerve.

The middle ear acts as a transformer by increasing sound pressure and decreasing volume velocity (see, e.g., von Helmholtz, 1877; Wever and Lawrence, 1954; Dallos, 1973; Shera and Zweig, 1992; Rosowski, 1994). Middle ear function has been characterized in several species by measuring frequency-dependent transfer functions between the sound pressure in the ear canal (the input to the middle ear) and the input to the cochlea. One such function is the stapes volume-velocity transfer function (**SVTF**)[1] defined as the ratio between stapes volume velocity $\mathbf{U_S}$ and ear canal sound pressure near the TM $\mathbf{P_{TM}}$ ($\mathbf{SVTF}=\mathbf{U_S}/\mathbf{P_{TM}}$). **SVTF** has been studied in many species, using a variety of techniques. Another transfer function of interest is the middle ear pressure gain ($\mathbf{G_{ME}}$) defined as the ratio between the pressure in the vestibule ($\mathbf{P_V}$) and $\mathbf{P_{TM}}$ ($\mathbf{G_{ME}}=\mathbf{P_V}/\mathbf{P_{TM}}$).

Because of the difficulties associated with intracochlear pressure measurements, $\mathbf{G_{ME}}$ has been reported in only a few individuals from a few species: in cat (Nedzelnitsky, 1980; Décory, 1989; Décory *et al.*, 1990), guinea pig (Dancer and Franke, 1980; Décory, 1989; Décory *et al.*, 1990; Magnan *et al.*, 1997), chinchilla (Décory, 1989; Décory *et al.*, 1990), gerbil (Olson, 1998, 2001; Dong and Olson, 2006; de La Rochefoucauld *et al.*, 2008) and human temporal bone (Puria *et al.*, 1997; Aibara *et al.*, 2001; Puria, 2003; Nakajima *et al.*, 2009). The difficulties in $\mathbf{P_V}$ measurements arise from the requirement of drilling a hole in the base of the cochlea to insert a pressure sensor. The hole needs to be small enough so that the middle and inner ear structures are not damaged, and the sensor needs to be small with a large impedance so as not to disturb the pressure waves being measured.

The cochlear input impedance $\mathbf{Z_C}$, the frequency-dependent complex ratio of $\mathbf{P_V}$ to stapes volume velocity $\mathbf{U_S}$ ($\mathbf{Z_C}=\mathbf{P_V}/\mathbf{U_S}$), quantifies the acoustic load of the inner ear on the middle ear and is important for our understanding of middle ear function. Changing this load can induce changes in middle ear input admittance $\mathbf{Y_{ME}}$ (the ratio of the TM volume velocity to $\mathbf{P_{TM}}$), stapes motion, and auditory sensitivity: For example, draining the cochlea introduced a resonance at 250 Hz in the middle ear input admittance of chinchillas and suppressed the non-linear dependence on level of $\mathbf{Y_{ME}}$ observed below 300 Hz (Rosowski *et al.*, 2006). The same manipulation in cats resulted in the introduction of various resonances and anti-resonances in $\mathbf{Y_{ME}}$ between 1 and 30 kHz (Peake *et al.*, 1992; Puria and Allen, 1998). In human temporal bones, complex rotary stapes motion, in addition to the piston component, was observed at high frequency when the cochlea was fluid-filled, but not after it was drained (Hato *et al.*, 2003). A less dramatic change in $\mathbf{Z_C}$ can also have an effect: In chinchillas, introducing an opening in the bony superior semi-circular canal, which reduces $\mathbf{Z_C}$ in a

frequency-dependent manner (Songer and Rosowski, 2007b), produced an increase in the magnitude of $\mathbf{Y_{ME}}$ between 100 Hz and 1 kHz and an increase in the magnitude of **SVTF** between 120 Hz and 5.8 kHz (Songer and Rosowski, 2006), as well as a broadband decrease in auditory sensitivity as measured by cochlear potentials (Songer and Rosowski, 2005).

Knowledge of $\mathbf{Z_C}$ is also important to the understanding of sound power flow through the middle and inner ears: The real part of $\mathbf{Z_C}$ can be used to compute the acoustic power entering the cochlea W_C (see Sec. IV G), where reasonable matches between auditory thresholds and the sound pressure necessary to produce a constant W_C in various species suggest middle ear power transfer is an important determinant to auditory sensitivity (Khanna and Tonndorf, 1969; Rosowski, 1991).

$\mathbf{Z_C}$ has been computed previously from simultaneous or non-simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ in several mammalian species. By definition, $\mathbf{U_S}$ rather than $\mathbf{V_S}$ is needed to compute $\mathbf{Z_C}$. Assuming piston-like motion of the stapes, $\mathbf{U_S}$ is deduced from $\mathbf{V_S}$ by multiplication of the stapes footplate area A_{FP}: $\mathbf{U_S}=A_{FP}\mathbf{V_S}$. Estimates of $\mathbf{Z_C}$ have been computed from the combination of non-simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ in guinea pig (Dancer and Franke, 1980), chinchilla (Ruggero *et al.*, 1990), human temporal bone (Puria *et al.*, 1997), and gerbil (Overstreet and Ruggero, 2002; Decraemer *et al.*, 2007). Simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ were performed in only a few individuals from cats (Lynch *et al.*, 1982) and gerbils (de La Rochefoucauld *et al.*, 2008), as well as in human temporal bone (Aibara *et al.*, 2001; Puria, 2003; Nakajima *et al.*, 2009). In these studies, estimates of $\mathbf{Z_C}$ are different than in studies combining non-simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$: In human temporal bone, estimates from simultaneous and non-simultaneous measurements showed differences in magnitude and phase (Puria *et al.*, 1997; Aibara *et al.*, 2001), but there were also important differences between the three studies using simultaneous measurements (Aibara *et al.*, 2001; Puria, 2003; Nakajima *et al.*, 2009). In gerbil, the magnitude of the non-simultaneous estimate (Overstreet and Ruggero, 2002) agreed with the simultaneous estimate (de La Rochefoucauld *et al.*, 2008), but there were differences in the phase, especially above 6 kHz. Moreover, an estimate of $\mathbf{Z_C}$ from simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ in one animal showed differences in its detailed structure from an estimate from non-simultaneous measurements in the same animal (de La Rochefoucauld *et al.*, 2008). Based on these comparisons, it seems important to measure $\mathbf{P_V}$ and $\mathbf{V_S}$ simultaneously to characterize $\mathbf{Z_C}$ more precisely.

We measured $\mathbf{P_V}$ and $\mathbf{V_S}$ simultaneously in chinchillas, a species whose hearing range (20 Hz–30 kHz) is similar to that of humans, using miniature fiber-optic pressure sensors and laser Doppler vibrometry in response to acoustic stimulation. Based on these measurements, we present a new estimate of $\mathbf{G_{ME}}$ in this species, extending previous measurements to higher frequencies, and, to our knowledge, the first estimate of $\mathbf{Z_C}$ in chinchillas derived from simultaneous measurements. $\mathbf{G_{ME}}$ had a magnitude of the order of 30 dB and a phase close to 0 between 300 Hz and 4 kHz, similar to the predictions of a simple transformer ratio model. Our $\mathbf{G_{ME}}$ was similar to published data in the chinchilla (Décory, 1989; Décory *et al.*, 1990) at stimulus frequencies of 500 Hz to 3 kHz and 12 to 20 kHz, but had a larger magnitude at other frequencies. $\mathbf{Z_C}$ had a flat magnitude and a phase close to 0, which are the characteristics of an acoustic resistance, over a broad frequency range. Our $\mathbf{Z_C}$ had a larger magnitude with a smaller slope below 1 kHz and a phase angle closer to 0 than a previous estimate from non-simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ (Ruggero *et al.*, 1990). We also compare measurements of auditory threshold in chinchilla (Miller, 1970) with estimates of the sound pressure required to produce a constant $\mathbf{P_V}$ across frequency, a constant $\mathbf{V_S}$, and a constant power into the vestibule. Overall, these new findings extend our knowledge of the chinchilla middle ear, and, together with other measured middle and inner ear mechano-acoustic quantities in this species, help draw a more complete picture of middle ear function. Understanding sound transfer through the middle ear could eventually improve therapeutic approaches to diseases and malfunctions of the middle ear.

II. MATERIALS AND METHODS

A. Fiber-optic pressure sensors

Measurements of $\mathbf{P_V}$ are constrained by the limited space available for the pressure sensors (the volume of the chinchilla vestibule is on the order of 5 μl: see Plontke *et al.*, 2002) and the fragility of the inner and middle ear structures. We chose to use fiber-optic pressure sensors (Olson, 1998) because they have good sensitivity, good high frequency response (up to about 100 kHz), and because they are very small. Their small size (about 145 μm in diameter) ensures the following:

(a) minimal disruption of the pressure field in the inner ear at frequencies in the chinchilla's hearing range, because of both the small size (145 μm is less than 1% of the wavelength in water at 30 kHz) and the relatively high impedance associated with such a small microphone, and
(b) minimal damage to the middle and inner ear structures during insertion into the vestibule.

The fiber-optic pressure sensors were fabricated following the techniques of Olson (1998). They are composed of a glass capillary tube (145 μm outer diameter) with a gold-coated polymer diaphragm affixed to one end. A single optical fiber (100 μm outer diameter) is inserted into the other end. The optical fiber is spliced to a *Y* coupling. A light emitting diode (LED) attached to one coupler branch produces incoherent light, and a photodiode attached to the other branch measures the light reflected from the diaphragm. Sound pressure flexes the diaphragm and modulates the reflected light.

Similarly to Olson (1998), calibration of the sensors was performed in water, according to the method described by Schloss and Strasberg (1962): The sensor is immersed in a column of liquid that is shaken vertically by a known amount (Brüel & Kjær Vibration Calibration Exciter Type 4290 with built-in reference accelerometer; Brüel & Kjær Measuring

Amplifier Type 2525); the pressure at the diaphragm depends on both the depth of immersion and on the acceleration of the column. Calibration functions were defined as the ratio between the sound pressure at the sensor diaphragm and the voltage at the output of the photodiode. For typical sensors, the calibration function magnitude was essentially flat up to 10 kHz, with a value of about 150 dB re 20 μPa/V, and then decreased between 10 and 30 kHz by about 8 dB. The phase was roughly flat and close to 0 on the entire range of measurements. In earlier experiments, we also calibrated the sensors in air, against a reference microphone of known sensitivity. In our experience, the two calibration methods (in water or in air) led to very similar sensitivities in magnitude and phase. Therefore, we only performed water calibration in subsequent experiments.

The main issues with the fiber-optic pressure sensors are their fragility and their stability. We periodically recalibrated the sensor to test for variations in its sensitivity. We report data only from sensors whose calibration was stable throughout the measurement session: In three animals, there was less than a 2 dB variation over the entire tested frequency range between repeated calibrations throughout the experiment; in two other animals, calibration was repeatable within 2 dB in the mid- and high frequencies, and within 5–8 dB in the lows; in another animal, calibration was repeatable within 10 dB. The calibration function used to compute $\mathbf{P_V}$ was the closest in time to the measurement.

Stability of the pressure sensor was also continuously monitored via the direct current (DC) component of the sensor photodiode output. In many sensors, a large change in the DC value was correlated with a change in sensitivity. Therefore, we report $\mathbf{P_V}$ measurements for which the DC value of the sensor during the measurements was close to the DC value during calibration: In the 6 animals from which data are presented, the change in DC between calibration and measurement was less than 8% (8% in one animal, 4% in two animals, 3% in one animal, 2% in one animal, and $<1\%$ in one animal).

B. Laser Doppler vibrometry

To measure stapes velocity, we used a single-beam laser Doppler vibrometer (Polytec CLV 700) aimed at small reflective plastic beads (<50 μm diameter) placed on the posterior crus and the footplate. Sound-induced velocity of the stapes was measured using the Doppler shift of light reflected from the moving beads. The sensitivity of the laser was checked by comparing the velocity of a shaker as measured by the laser with its acceleration as measured by the reference accelerometer. The sensitivity was constant over the frequency range of measurement.

Our surgical exposure of the stapes allowed measurement of the piston-like component of stapes motion within an angle of about 30° relative to the piston-like direction. The volume velocity $\mathbf{U_S}$ was estimated assuming piston-like motion of the stapes. In this case, the volume velocity $\mathbf{U_S}$ is simply the product of measured linear velocity $\mathbf{V_S}$ and the average area of the chinchilla footplate (A_{FP}=2 mm^2, Vrettakos *et al.*, 1988).

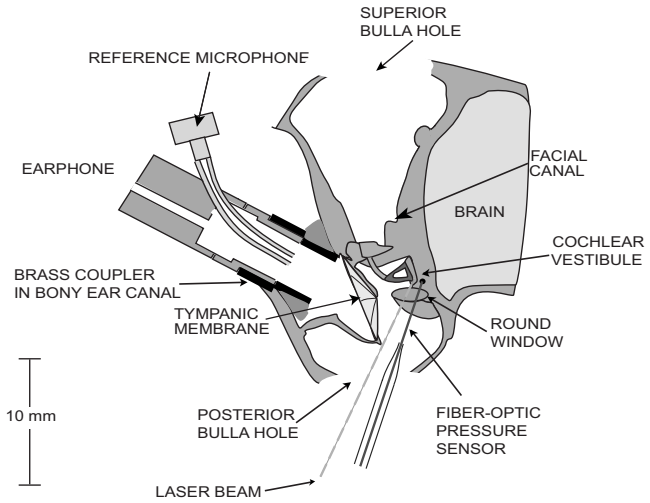


FIG. 1. Schematic of the animal preparation. The bulla was opened superiorly and posteriorly, and a hole approximately 200 μm in diameter was made in the vestibule for the fiber-optic pressure sensor. The cartilaginous ear canal was cut and a brass tube was glued in the bony ear canal to allow repeatable couplings of the earphone delivering the sound stimuli. A built-in reference microphone measured sound pressure in the ear canal. A laser Doppler vibrometer was aimed at reflective beads placed on the posterior crus and the footplate of the stapes.

C. Animal preparation

The surgeries and experiments were performed in accordance with guidelines published by the U.S. Public Health Service and were approved by the Massachusetts Eye and Ear Infirmary (MEEI) Animal Care Committee.

The animals were anesthetized with intraperitoneal injections of sodium pentobarbital (50 mg/kg) and Ketamine (40 mg/kg). After a tracheotomy to facilitate respiration, an opening was made in the superior bulla. The tensor tympani muscle and the tympanic branch of the facial nerve, which innervates the stapedius muscle, were cut to prevent random contractions of these muscles during the experiment (Rosowski *et al.*, 2006). A second hole in the posterior bulla was made to view the stapes and round window. Part of a bony wall posterior to the round window, in which the facial nerve passes, was removed in order to see the wall of the vestibule posterior to the stapes. In doing so, extreme care was taken to avoid pulling or damaging the stapedius tendon. A hole for placement of the $\mathbf{P_V}$ fiber-optic pressure sensor, of approximate diameter 180–250 μm, was made in the vestibule 1–2 mm away from the oval window with a fine sharp pick (Fig. 1). On several occasions, we verified that this hole opened into the vestibule by breaking the annular ligament and pushing the stapes into the vestibule; the footplate was then seen through the $\mathbf{P_V}$ hole.

The cartilaginous ear canal was cut and a brass tube was placed and glued in the bony ear canal to allow repeatable couplings of the earphone delivering the sound stimuli. The middle ear was open during the measurements.

D. Stimuli and responses

In early experiments we used an electrodynamic hearing-aid earphone (Knowles, Itasca, IL) coupled to a power amplifier as our sound source. In later experiments, an

electrostatic earphone (Tucker-Davis Technologies Type EC1, driven by a Tucker-Davis Technologies ED1 earphone driver) was used. The selected speaker was sealed to the brass tube in the ear canal, and the input of the speaker driver was coupled to the output of a programmable attenuator (Tucker-Davis Technologies Type PA5). We used LABVIEW (National Instruments, Austin, TX) to construct stimuli, control the attenuator, and measure the voltage outputs of our different sensors. The stimuli consisted of stepped pure tones (six points/octave) from 62.5 Hz to 28.5 kHz, with sound pressure levels (SPLs) ranging from 65 to 110 dB SPL at the entrance of the brass tube. The duration of each tone was 1–2 s; the responses were averaged and Fourier analysis was used to quantify the magnitude and phase of the fundamental and harmonic components of the response as well as the stimulus noise floor.

E. Correction of ear canal pressure measurements

Although the middle ear pressure gain $\mathbf{G_{ME}}$ is defined as the ratio between $\mathbf{P_V}$ and $\mathbf{P_{TM}}$ (the sound pressure at the TM near the umbo), during most of our measurements, we measured ear canal sound pressure ($\mathbf{P_{EC}}$) using a reference microphone built into the entrance of the brass coupling tube, about 7 mm from the umbo (Fig. 1). At frequencies less than 5 kHz (where the wavelength of sound is about 70 mm), the sound pressure in the ear canal and the sound pressure near the TM are essentially equal ($\mathbf{P_{TM}} \sim \mathbf{P_{EC}}$), but at higher frequencies significant differences occur between $\mathbf{P_{TM}}$ and $\mathbf{P_{EC}}$. To account for these differences, we used an average correction computed from simultaneous measurements of $\mathbf{P_{TM}}$ and $\mathbf{P_{EC}}$ in nine ears used in other studies (included in Ravicz *et al.*, 2009) to compute a correction factor $\mathbf{P_{TM}}/\mathbf{P_{EC}}$ that was used to correct our measurements of $\mathbf{G_{ME}}$ and **SVTF** (see Sec. IV). No correction was necessary for $\mathbf{Z_C}$, whose computation from simultaneously measured $\mathbf{P_V}$ and $\mathbf{V_S}$ does not involve $\mathbf{P_{TM}}$.

F. Frequency range and choice of earphone

We estimated the signal-to-noise ratios (SNRs) for all of our measured quantities ($\mathbf{P_V}$, $\mathbf{V_S}$, and $\mathbf{P_{EC}}$) by comparing the magnitude of the response spectrum at the stimulus frequency to the magnitude at nearby frequencies that were not in the stimulus and not harmonics of the stimulus frequencies. The electrodynamic earphone used in earlier experiments had low outputs at frequencies above 15 kHz, which led to low signal levels and low SNRs in our measurements at higher frequencies. In later experiments, we used an electrostatic earphone and obtained a good SNR over most of the stimulus frequency range (out to 28.5 kHz).

We include only data for which the SNR was larger than 10 dB. Moreover, in all the plots presented in this paper, we marked with a circle all the mean data points for which the SNR of the measurements involved was less than 20 dB in at least one of the animals. We further tested the reliability of the measurements by testing their repeatability: In most cases, $\mathbf{G_{ME}}$ and **SVTF** were nearly exactly repeatable when the SNR was greater than 10 dB in all channels, but in some cases, we further restricted the frequency range to exclude data points that lacked repeatability.

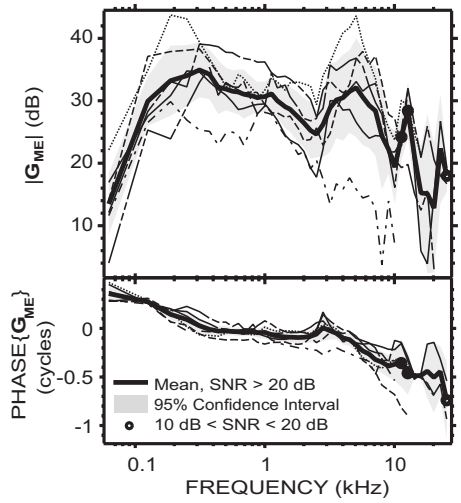


FIG. 2. $\mathbf{G_{ME}}$ in six animals (individual data: thin lines; mean: thick solid line; 95% CI: shaded areas). Data points with a SNR < 10 dB were not included; frequencies for which SNR < 20 dB in at least one of the individual data point were marked by a circle (see Sec. II).

G. Statistics

The mean and standard deviation of $\mathbf{G_{ME}}$'s decibel magnitude was computed from the decibel magnitudes of each animal. For **SVTF** and $\mathbf{Z_C}$, we used the mean and standard deviation of the logarithm of the magnitudes, and took their exponential to go back to the linear domain. For the phases of those three functions, we computed the means and standard deviations of the unwrapped individual phases. In all cases, we included only the data points that passed our SNR and repeatability criteria in the computation of the mean and standard deviation. Therefore, the number of animals included depended on frequency and was, in particular, lower at high frequencies.

In our plots, we usually show the mean and 95% confidence interval (CI). The 95% CI is defined by two standard errors of the mean above and below the mean. We computed the standard error by dividing the standard deviation by the square root of the frequency-dependent number of animals.

III. RESULTS

Thirteen animals were used in this study. Among these, four had their middle or inner ears damaged during surgery. In three other experiments the pressure sensor proved unstable. We therefore present $\mathbf{G_{ME}}$, **SVTF**, and $\mathbf{Z_C}$ results in six animals. For each experiment, we restrict the results shown to the frequency range over which the measurements were repeatable and had a SNR larger than 10 dB (see Sec. II).

A. Middle ear pressure gain $\mathbf{G_{ME}}$

$\mathbf{G_{ME}} = \mathbf{P_V}/\mathbf{P_{TM}}$ was computed from simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{P_{EC}}$, and corrected for the differences between $\mathbf{P_{TM}}$ and $\mathbf{P_{EC}}$, as explained in Sec. II. $\mathbf{G_{ME}}$ in each animal is plotted in Fig. 2. Both magnitude and phase angle

were similar among the six animals. The 95% CI was between 2 and 8 dB for the magnitude over almost the entire frequency range of measurement, and less than 0.1 cycles for the phase below 8 kHz. One animal was an outlier with low gain magnitude and more negative phase angles at high frequencies. The mean $|\mathbf{G_{ME}}|$ across these six ears ranged from 13 to 35 dB over the measurement frequencies. Between 62 and 125 Hz, the mean $|\mathbf{G_{ME}}|$ rapidly increased from 14 dB to 30 dB, with a slope of 16 dB/octave. It then continued to increase with a slower 4 dB/octave slope to reach a maximum of 35 dB at 375 Hz. Between 375 Hz and 2.5 kHz, $|\mathbf{G_{ME}}|$ slowly decreased to 25 dB with an average slope of −3.7 dB/octave. It then increased sharply to reach 32 dB at 5 kHz (10 dB/octave slope), and decreased sharply to reach a 19 dB local minimum at 10 kHz (−13 dB/octave slope). Above 10 kHz, the mean $|\mathbf{G_{ME}}|$ displays a 28 dB peak at 12.7 kHz, a 13 dB notch at 20.2 kHz, and a 22 dB peak at 22.6 kHz.

$\mathbf{G_{ME}}$'s mean phase angle (which is the phase difference between $\mathbf{P_V}$ and $\mathbf{P_{TM}}$) decreased from 0.4 to 0 cycles from 62 to 300 Hz, was near 0 between 0.3 and 3 kHz, and accumulated with frequency above that, reaching −0.4 cycles by 10 kHz and −0.7 cycles by 25 kHz. The ripple near 3 kHz associated with a dip in $|\mathbf{G_{ME}}|$ is likely due to a resonance between the bulla air space and the open holes in the bullar walls (Rosowski *et al.*, 2006). The group delay[2] was about 45 μs between 500 Hz and 2 kHz and about 61 μs between 3 and 9 kHz.

von Helmholtz (1877) proposed that the middle ear works as an acoustico-mechanical transformer composed of several levers in cascade, to increase sound pressure from the ear canal to the cochlea. It is interesting to note that our mean $|\mathbf{G_{ME}}|$ reached a maximum comparable to this "transformer ratio" of 38 dB computed from anatomical data (Vrettakos *et al.*, 1988)[3] at 300 Hz (see Fig. 2) and was within 10 dB of this value over a wide frequency range (roughly 125 Hz–8 kHz). Moreover, the phase was close to 0 (within 0.1 cycles) in a similar frequency range (roughly 250 Hz–4 kHz), which is also consistent with an ideal transformer at these frequencies. However, outside this frequency range, the magnitude and phase characteristics largely differed from that of an ideal transformer, confirming that the ideal transformer hypothesis does not accurately describe middle ear function at all frequencies (e.g., Funnell, 1996).

We tested the stability of the preparation and the pressure measurement sensor by taking the pressure sensor out of the inner ear, recalibrating, reinserting the sensor into the vestibule, and remeasuring $\mathbf{P_V}$. Figure 3(A) compares the initial and repeated measurements in an experiment where the pressure sensor was very stable (recalibration sensitivity was within 2 dB of the original sensitivity estimate, and the photodiode DC output was constant). $|\mathbf{G_{ME}}|$ upon reinsertion was within 2 dB of the original measurement over the entire range of frequency. The phase angles were also almost identical, with variations of less than 0.01 cycles at most test frequencies. These results are consistent with a highly stable preparation.

In some experiments we also verified that the gain we measured was indeed due to the middle ear and not some other artifactual signal path (e.g., bone conduction) by measuring $\mathbf{G_{ME}}$ before and after interrupting the incudo-stapedial joint. Figure 3(B) shows an example from one experiment. Before the interruption, $\mathbf{G_{ME}}$ had the characteristic magnitude and phase described earlier. After interrupting the incudo-stapedial joint (and making sure that the lenticular process was not touching the stapes head), $|\mathbf{G_{ME}}|$ dropped dramatically at all frequencies, essentially reaching the noise floor of our sensor (about 50 dB below the pressure measured with an intact joint, except as shown). This control is a clear demonstration that we measured ossicular sound conduction.

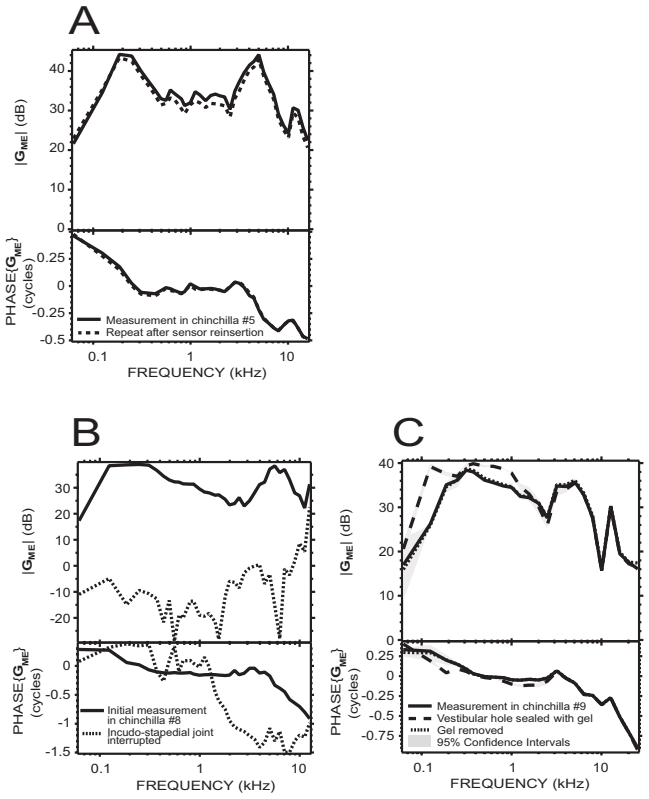


FIG. 3. $\mathbf{G_{ME}}$ controls. (A) Stability: comparison of $\mathbf{G_{ME}}$ measured in chinchilla No. 5 before (solid line) and after (dotted line) taking the pressure sensor out of the vestibule and recalibrating it. Differences were less than 2 dB in magnitude and less than 0.01 cycles in phase. (B) Effect of interrupting the ossicular chain: $\mathbf{G_{ME}}$ in chinchilla No. 8 with an intact ossicular chain (solid line), and after interrupting the incudo-stapedial joint (dotted line). (C) Effect of plugging the vestibular hole on $\mathbf{G_{ME}}$: in one animal (chinchilla No. 9), after sealing around the pressure sensor with a gel of high molecular weight (dashed line), there was a limited increase in $|\mathbf{G_{ME}}|$. After removing the gel (dotted line), $\mathbf{G_{ME}}$ went back to baseline (solid line). The 95% CI (shaded areas) represent the variability of repeated measurements in the same condition.

To determine the influence of the vestibular hole with the inserted pressure sensor on $\mathbf{G_{ME}}$, we tried to seal the pressure sensor in place with dental impression material (Jeltrate), dental cement, or a sodium hyaluronate viscoelastic gel of high molecular weight (Healon GV 14 mg/ml). With a hole of approximate diameter 180–250 μm and a 145 μm diameter sensor, the area of the hole not occupied by the sensor was roughly 0.009–0.026 mm^2. In most preparations, it was not possible to seal around the sensor effectively because of the limited space available and because the outward flow of perilymph pushed the sealant material away. In one

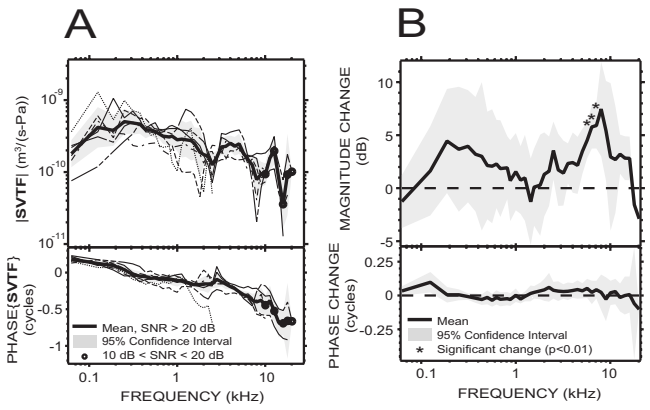


FIG. 4. **SVTF** in six animals. (A) **SVTF** before the vestibular hole was made (individual data: thin lines; mean: thick solid line; 95% CI: shaded areas). (B) Mean effect of introducing the vestibular hole on **SVTF** in six animals. We computed the change in magnitude and phase before the hole was made and afterward with the pressure sensor in place. Changes in magnitude were significant ($p<0.01$) only in a small region around 6 kHz (asterisks).

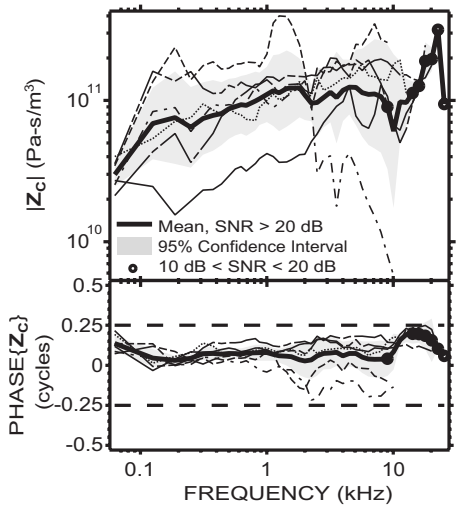


FIG. 5. $\mathbf{Z_C}$ in six animals (individual data: thin lines; mean: thick solid line; 95% CI: shaded areas) from simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$. $\mathbf{Z_C}$ was resistive (roughly flat magnitude, phase close to 0) up to at least 10 kHz.

case, as shown in Fig. 3(C), the viscoelastic gel appeared to cover most of the hole, resulting in an increase in $\mathbf{P_V}$, and therefore in $\mathbf{G_{ME}}$, especially at frequencies below 1 kHz: $|\mathbf{G_{ME}}|$ increased by as much as 15 dB at 125 Hz and the phase decreased by about 0.15 cycles at 200 Hz. After removing the gel, $\mathbf{P_V}$ went back to the lower level, and $\mathbf{G_{ME}}$ was very similar in amplitude and phase to the initial measurement. Several other attempts at sealing the hole produced smaller changes. Overall, the effect of sealing the hole on $\mathbf{G_{ME}}$ was small and limited to the lowest measured frequencies.

B. Stapes velocity transfer function SVTF

$\mathbf{SVTF}=\mathbf{U_S}/\mathbf{P_{TM}}$ was computed in each animal from the measured stapes velocity $\mathbf{V_S}$ and a 2 mm^2 mean stapes footplate area as described in Sec. II, normalized by the measured $\mathbf{P_{EC}}$, and corrected by the $\mathbf{P_{TM}}/\mathbf{P_{EC}}$ transfer function. $\mathbf{V_S}$ was measured in two conditions:

(1) with an intact vestibule, and
(2) after the vestibular hole was drilled and the pressure sensor inserted in the vestibule.

In the intact vestibule condition [Fig. 4(A)], **SVTF** was very similar among the six ears, except for one phase outlier (the outlier is terminated at 2.5 kHz due to SNR problems). The range of the 95% CI was within a factor of 2–3 for the magnitude, and less than 0.1 cycles for the phase angle (equal to the phase difference between $\mathbf{V_S}$ and $\mathbf{P_{TM}}$), over most of the frequency range of measurement. $|\mathbf{SVTF}|$ increased by about a factor of 2.8 with frequency over the two-octave range of 62–250 Hz, and the phase decreased from +0.25 to 0 cycles over the same frequency range. The phase of the admittance (inverse of an impedance) of a combination of compliance and resistance is between 0 and +0.25 cycles. **SVTF** is an admittance as it is the ratio of a volume velocity ($\mathbf{U_S}$) to a pressure ($\mathbf{P_{TM}}$); therefore the phase over 62–250 Hz is consistent with a compliance-resistance combination. $|\mathbf{SVTF}|$ decreased by a factor of 3.4 over the three octaves between 0.3 and 2.4 kHz and the phase was between 0 and −0.25 cycles, consistent with a mass-resistance combination. $|\mathbf{SVTF}|$ increased by a factor of 1.7 between 2.5 and 5 kHz, and decreased by a factor of 3 between 5 and 8 kHz, while the phase decreased toward −0.4 cycles. The notch in $|\mathbf{SVTF}|$ associated with a ripple in the phase near 3 kHz is similar to what was observed in $\mathbf{G_{ME}}$ and is consistent with a resonance of the compliant air within the bulla and the acoustic mass associated with the holes in the bullar wall. Between 8 and 20 kHz, $|\mathbf{SVTF}|$ is characterized by a peak centered at 12.7 kHz and a notch centered at 16 kHz; the phase further decreased, reaching −0.7 cycles by 14 kHz, and then flattened. The group delay was about 51 μs between 500 Hz and 2 kHz and about 44 μs between 3 and 9 kHz.

To assess the influence of the hole in the vestibule wall on **SVTF**, we compared measurements of **SVTF** before the hole was made and afterward with the pressure sensor in place. We found small (<7 dB) increases in $|\mathbf{SVTF}|$ in the condition with the vestibular hole [Fig. 4(B)], which is consistent with the hole decreasing cochlear input impedance and facilitating stapes motion. A paired t-test performed at each frequency showed that the changes were significant ($p<0.01$) only in a small region around 6 kHz.

C. Cochlear input impedance $\mathbf{Z_C}$

$\mathbf{Z_C}=\mathbf{P_V}/\mathbf{U_S}$ was computed from simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$. As mentioned earlier, the computation of $\mathbf{Z_C}$ does not use $\mathbf{P_{TM}}$; therefore our results are not affected by the correction employed to convert $\mathbf{P_{EC}}$ to $\mathbf{P_{TM}}$. $\mathbf{Z_C}$ was similar among six ears (Fig. 5), except for a low outlier for $|\mathbf{Z_C}|$ in one ear at low frequencies and a low outlier in a different ear at high frequencies (corresponding to the ear with a low $|\mathbf{G_{ME}}|$ at high frequencies).

The average $|\mathbf{Z_C}|$ was about 10^{11} acoustic ohms. The factor of 3 range of the 95% CI is similar to the range of 95% CI in $\mathbf{G_{ME}}$ and **SVTF**. Between 62 and 125 Hz, $|\mathbf{Z_C}|$ in-

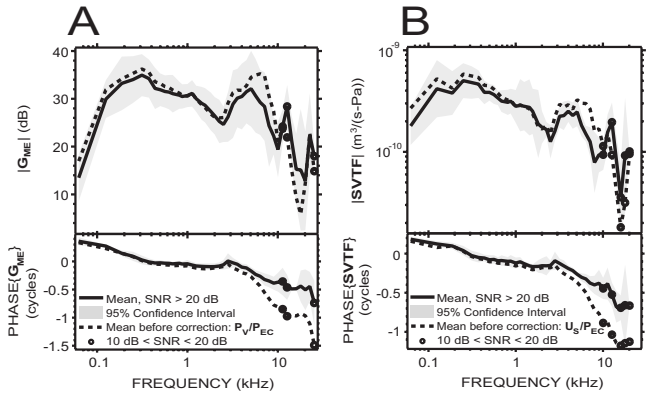


FIG. 6. Effect of an ear canal correction (to account for the differences between $\mathbf{P_{EC}}$ and $\mathbf{P_{TM}}$—see Sec. II) on $\mathbf{G_{ME}}$ and **SVTF**. (A) Mean $\mathbf{G_{ME}}$ before (dotted line) and after (solid line, from Fig. 2) correction. (B) Mean **SVTF** in the intact vestibule condition before (dotted line) and after [solid line, from Fig. 4(A)] correction.

creased and the phase (equal to the phase difference between $\mathbf{P_V}$ and $\mathbf{U_S}$) decreased from close to 0.125 cycles to about 0; this is consistent with a mass-resistance combination (phase between 0 and +0.25 cycles) at the lowest frequency tested. $|\mathbf{Z_C}|$ was roughly constant with frequency between 125 Hz and 10 kHz, and the phase was near zero in this frequency range; this is consistent with a resistance. $|\mathbf{Z_C}|$ increased sharply from 10 to 22 kHz and fell sharply at 25 kHz. Consistent with the input impedance of a passive system, the phase had values between −0.25 and +0.25 cycles at all frequencies measured.

IV. DISCUSSION

We simultaneously measured the sound pressure in the vestibule $\mathbf{P_V}$ and stapes velocity $\mathbf{V_S}$ in chinchilla preparations in response to acoustic stimulations. The middle ear gain $\mathbf{G_{ME}}$, stapes velocity transfer function **SVTF**, and cochlear input impedance $\mathbf{Z_C}$ were computed from these measurements in six ears where the middle and inner ear structures were intact and the pressure sensors stable. In this section, we (a) discuss the role of the ear canal correction we applied to **SVTF** and $\mathbf{G_{ME}}$; (b) address some other factors that may bias our high frequency responses; (c) assess the influence of the vestibular hole on our estimates in the context of (d) an acoustic model; (e) compare our results with other studies in chinchilla and (f) other mammals; and (g) test the hypothesis that auditory thresholds can be predicted by $\mathbf{P_V}$, $\mathbf{U_S}$, or the acoustic power delivered to the cochlea.

A. Effect of ear canal correction on $\mathbf{G_{ME}}$ and SVTF

The $\mathbf{G_{ME}}$ and **SVTF** data presented in Figs. 2 and 4(A), respectively, have been corrected by the mean ratio of sound pressures measured near the umbo ($\mathbf{P_{TM}}$) and about 7 mm from the umbo ($\mathbf{P_{EC}}$) measured in nine other ears (a location that corresponded to the $\mathbf{P_{EC}}$ microphone in these experiments), as described in Sec. II. The effects of this $\mathbf{P_{TM}}/\mathbf{P_{EC}}$ correction function on $\mathbf{G_{ME}}$ and **SVTF** are shown in Fig. 6. As expected, the effect was small below 3 kHz, because the wavelength of sound at these low frequencies is much longer than the dimensions of the ear canal and the sound pressure is approximately constant throughout the ear canal (e.g., Beranek, 1986). Above 3 kHz, the correction caused significant changes in both $\mathbf{G_{ME}}$ and **SVTF** and reduced the measured group delay of both $\mathbf{G_{ME}}$ and **SVTF** by about 45 μs.

Above 3 kHz, the corrected and uncorrected mean $|\mathbf{G_{ME}}|$s [Fig. 6(A)] are similar in shape, but the correction had the effect of (a) shifting the 7 kHz maximum to 5 kHz, while decreasing $|\mathbf{G_{ME}}|$ by 4 dB on average between 2 and 10 kHz, (b) shifting the 11.5 kHz maximum to 12.5 kHz, while increasing $|\mathbf{G_{ME}}|$ by 4 dB between 10 and 20 kHz, and (c) shifting the 18 kHz minimum to 20 kHz, while increasing $|\mathbf{G_{ME}}|$ by about 7 dB. The effects of the correction on |**SVTF**| [Fig. 6(B)] above 3 kHz were to enhance the maximum at 5 kHz and the minimum at 16 kHz. High frequency correction effects are discussed further below.

B. High frequency responses

The high frequency responses we obtained for $\mathbf{G_{ME}}$ and **SVTF** are characterized by an increased variance in magnitude and/or phase relative to lower frequencies. For $|\mathbf{G_{ME}}|$, the 95% CI was 7–12 dB above 16 kHz, and generally 3–6 dB at lower frequencies (Fig. 2). Similarly for the phase, the 95% CI was greater than 0.1 cycles above 10 kHz, smaller below. As for **SVTF**, the 95% CI of the magnitude was roughly similar across frequencies, but the 95% CI of the phase was usually higher above 9 kHz [Fig. 4(A)]. These increased confidence intervals at high frequencies may be due to several factors.

(i) Differences between ear canal sound pressure at our measurement location $\mathbf{P_{EC}}$ and near the TM $\mathbf{P_{TM}}$: The correction factor $\mathbf{P_{TM}}/\mathbf{P_{EC}}$ we used to account for these differences was averaged over measurements in nine ears that were different from the ears in which we made the $\mathbf{P_V}$ and $\mathbf{V_S}$ measurements reported here. The peaks and valleys observed in $|\mathbf{P_{TM}}/\mathbf{P_{EC}}|$, in particular, above 10 kHz, are dependent on the anatomy of the ear canal, which varies among individuals. Consequently, we can expect the correction to be imperfect.

(ii) Fewer animals were included: In earlier experiments, the earphone we used did not provide sufficient sound pressure stimulus at frequencies above 15 kHz, so there are fewer data at high frequencies with a SNR greater than 10 dB (our criteria: see Sec. II). In particular, for $\mathbf{G_{ME}}$, data from six animals were included up to 10 kHz, but only two animals met the SNR criterion above 22 kHz; for **SVTF**, six animals were included up to 9 kHz, two are included at 20 kHz, and only one above that. The breakdown in number of animals included in the estimates of $\mathbf{Z_C}$ is similar to that for **SVTF**.

(iii) The signal-to-noise ratios in the included animals were lower: The data points for which the SNR is between 10 and 20 dB are all concentrated in the high frequencies (see the circles in Figs. 2, 4(A), and 5). Even though a 10 dB SNR means that the signal was

more than three times higher than the noise, noise may have contributed to the increased variance at high frequencies.

(iv) For $\mathbf{U_S}$, we assume piston-like motion of the stapes, and measure velocity in only one direction. During an experiment, the laser angle was set so as to have a clear view of reflectors on the stapes footplate or posterior crus. The actual measurement angle was therefore highly dependent on the animal's specific anatomy and position of the reflectors, usually within $30 \pm 10°$. If the motion is truly piston-like, our measurements are not very sensitive to the laser beam angle. For example, measuring with a 30° angle relative to the piston axis introduces an error corresponding to a factor of cos(30°) or −1.2 dB in measuring the piston component, and small variations about that angle only produce small changes in estimated piston-like velocity: A 40° angle leads to an error in the estimate of piston-like velocity of −2.3 dB, while a 20° angle results in a −0.5 dB error. Therefore, if stapes motion was truly piston-like, the uncertainty on the measurement angle would only result in less than a 3 dB error for these angle values.

Piston-like motion of the stapes is probably a valid assumption at low frequencies; however, it is unlikely that this assumption is valid at high frequencies, and other modes that cause transverse and rotational motion of the stapes at the measurement points may make large contributions to the measured $\mathbf{V_S}$ (e.g., in human temporal bone: Heiland *et al.*, 1999; Hato *et al.*, 2003; Chien *et al.*, 2006; in gerbil: Decraemer *et al.*, 2007; Ravicz *et al.*, 2008; and in cat and cat temporal bone: Guinan and Peake, 1967; Decraemer *et al.*, 2003; Decraemer and Khanna, 2003). The existence and variability of such non-piston modes of motion at high frequency could explain the larger variance for $\mathbf{U_S}$ and **SVTF** in that frequency range.

The 12 kHz peak and the 16 kHz notch we found in $|\mathbf{SVTF}|$ [Fig. 4(A)], as well as the sharp peak in $|\mathbf{Z_C}|$ at high frequencies (Fig. 5), may also be explained by complex motion of the stapes at high frequency. A hypothesis is that the rocking component of stapes motion above 12 kHz is significant, and would result in a measured linear motion of greater or smaller amplitude than the actual $\mathbf{V_S}$, depending on the relative phase between the translational and rocking movements. This hypothesis is credible in light of the aforementioned studies in human temporal bone, gerbil, and cat; however, a detailed description of non-piston modes for the chinchilla has not been done.

C. Influence of the vestibular hole

It was necessary to make a hole in the vestibule to introduce the pressure sensor and measure $\mathbf{P_V}$. Overall, the effects of the vestibular hole on $\mathbf{G_{ME}}$ and **SVTF** were small and limited in frequency: Sealing the hole in one ear produced a small increase in $|\mathbf{G_{ME}}|$ below 1 kHz [Fig. 3(C)], and introducing the hole produced an average increase in $|\mathbf{SVTF}|$ of less than 7 dB in 6 ears [Fig. 4(B)], which was not statistically significant ($p < 0.01$) at most measured frequencies.

The changes we observed for **SVTF** were consistent with a study by Songer and Rosowski (2006). In that study, they looked at the effect of semi-circular canal dehiscence on **SVTF** in chinchillas. They found an increase in $|\mathbf{SVTF}|$ above 100 Hz that was maximal at frequencies between 150 and 500 Hz (5–10 dB), decreased with frequency between 500 and 1000 Hz to a value of roughly 2 dB, and persisted at about 2 dB above the intact case from 1 to 7 kHz. It could not be determined if the increase in $|\mathbf{SVTF}|$ persisted to higher frequencies because measurements were noisy above 7 kHz. In our study, the increase in $|\mathbf{SVTF}|$ had a similar shape below 1 kHz, but had a lower value (the maximum of the mean was less than 5 dB at these frequencies) and this lower value was not statistically significant. We also found a small 0.1-cycle increase in phase angle between 100 and 200 Hz, smaller but consistent with Songer and Rosowski's (2006) 0.4 cycle increase at similar frequencies. The significant changes we observed in $|\mathbf{SVTF}|$ around 6–7 kHz are not visible in Songer and Rosowski's study, but this could be because of their noise issue at these high frequencies. There are also differences in the experimental setup: We introduced a small (180–250 μm diameter) hole, partially plugged (by a 145 μm diameter pressure sensor), in the wall of the vestibule, whereas they introduced a larger (500 μm diameter) open hole in the narrow superior semi-circular canal about 3–5 mm from the vestibule. The smaller changes we observed at low frequencies (5 dB on average in our case, and ~10 dB in their study) are consistent with the introduction of a smaller hole, while the lack of change at higher frequencies in the Songer and Rosowski data may be explained by the 3–5 mm narrow tube that separated their semi-circular canal dehiscence from the vestibule.

Changes in $\mathbf{G_{ME}}$ can be due to changes in $\mathbf{P_V}$, $\mathbf{P_{EC}}$, or both; similarly, changes in **SVTF** can be due to changes in $\mathbf{V_S}$, $\mathbf{P_{EC}}$, or both. It was not possible to ascertain accurately whether part of the changes we measured in $\mathbf{G_{ME}}$ and **SVTF** was due to changes in $\mathbf{P_{EC}}$: Comparison of the measured $\mathbf{P_{EC}}$ before the hole is made and afterward is not valid because we had to move the animal's head to make the hole, which sometimes changed the quality of the seal of the ear-phone in the brass-tube coupler, and affected $\mathbf{P_{EC}}$. Nevertheless, in three experiments, $\mathbf{P_{EC}}$ was nearly identical after the hole was introduced. It is highly unlikely that a change in the ear canal coupler seal would perfectly counteract an effect of the hole on $\mathbf{P_{EC}}$. There is a greater likelihood that in those experiments, the coupler seal did not vary after the hole introduction, and the changes we observed in $\mathbf{G_{ME}}$ and **SVTF** were mostly due to actual changes in $\mathbf{P_V}$ and $\mathbf{U_S}$.

D. Predictions of an acoustic model of a hole in the vestibule

Direct measurement of the effect of the vestibular hole on $\mathbf{G_{ME}}$, by plugging the hole around the pressure sensor, was difficult because of the proximity of the stapes, the flow of lymph out of the hole, and other anatomical and positional constraints. As we have controlled data for the effect of such

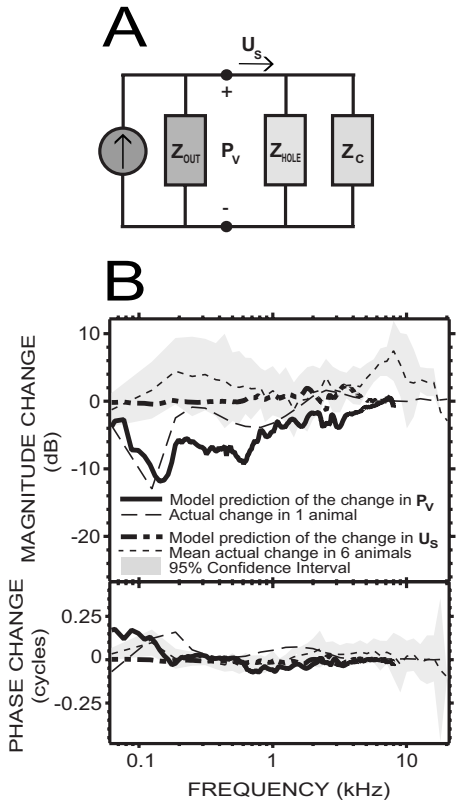


FIG. 7. Model predictions of the effect of the vestibular hole. (A) Simple model of the middle ear as a Norton equivalent circuit, providing volume velocity $\mathbf{U_S}$ to the parallel combination of the inner ear load ($\mathbf{Z_C}$) and the impedance of the hole ($\mathbf{Z_{hole}}$). The Norton equivalent is composed of an ideal volume-velocity source and the output impedance of the middle ear ($\mathbf{Z_{out}}$). (B) Predictions of our simple model with a 180 μm diameter hole partially plugged by a pressure sensor (equivalent radius=53 μm) were at least qualitatively similar to the experimental changes we found in $\mathbf{P_V}$ and $\mathbf{U_S}$ [see text and Figs. 3(C) and 4(B)]: The changes in $\mathbf{P_V}$ were limited to low frequencies (solid and dashed-dotted lines); the changes in $\mathbf{U_S}$ were small (dashed and dotted lines).

plugging on $\mathbf{G_{ME}}$ in only one animal, we now describe a lumped-element acoustic model that provides further insight into the influence of the hole.

The model we use to investigate the effect of the open hole around our $\mathbf{P_V}$ sensor [Fig. 7(A)] represents the middle ear as a Norton equivalent circuit (made up of an equivalent volume-velocity source and a parallel impedance equal to the output impedance of the middle ear and sound source $\mathbf{Z_{out}}$) that provides the stapes volume velocity $\mathbf{U_S}$ to the parallel combination of the inner ear load ($\mathbf{Z_C}$) and the impedance of the hole ($\mathbf{Z_{hole}}$). The pressure across each of the parallel branches of the circuit in Fig. 7(A) is $\mathbf{P_V}$. The changes in $\mathbf{P_V}$ and $\mathbf{U_S}$ introduced by opening the hole can be inferred from this circuit by simple linear current divider equations. If we further assume the output of the source $\mathbf{P_{EC}}$ is not affected by the introduction of the hole, as suggested by some of our experiments (see Sec. IV C above), we can define

$$\frac{\mathbf{P_{V,hole}}}{\mathbf{P_{V,normal}}} = 1 - \frac{\mathbf{Z_C Z_{out}}}{\mathbf{Z_C Z_{out}} + \mathbf{Z_C Z_{hole}} + \mathbf{Z_{hole} Z_{out}}}, \tag{1a}$$

$$\frac{\mathbf{U_{S,hole}}}{\mathbf{U_{S,normal}}} = 1 + \frac{\mathbf{Z_C}^2}{\mathbf{Z_C Z_{out}} + \mathbf{Z_C Z_{hole}} + \mathbf{Z_{hole} Z_{out}}}, \tag{1b}$$

where these ratios represent the change in $\mathbf{G_{ME}}$ and **SVTF**.

These ratios depend on three unknown impedances: $\mathbf{Z_C}$, $\mathbf{Z_{out}}$, and $\mathbf{Z_{hole}}$. We used estimates of $\mathbf{Z_C}$ and $\mathbf{Z_{out}}$ by Songer and Rosowski (2007a), which they computed based on a transmission matrix model of the middle ear, fed by measurements of ear canal pressure and stapes velocity in chinchillas below 8 kHz.[4] To compute $\mathbf{Z_{hole}}$, we modeled the hole as a lossy transmission line.[5] This model was originally developed by Egolf (1977), and used to model fluid-filled tube segments by Songer and Rosowski (2007b). In our case, $\mathbf{Z_{hole}}$ is computed as follows:

$$\mathbf{Z_{hole}} = \frac{\mathbf{A z_0} + \mathbf{B}}{\mathbf{C z_0} + \mathbf{D}}, \tag{2}$$

with $\mathbf{z_0}$ the termination impedance of the hole, and **A**, **B**, **C**, and **D** parameters depending on various thermodynamic parameters of the medium, frequency, and the dimensions of the hole. A detailed description of these parameters can be found in Songer and Rosowski (2007b).

The original model is for a tube of radius a and length l. For our purpose, this description is not entirely satisfying because the hole is partially obstructed by the pressure sensor. In order to apply the model, we computed an "equivalent radius" corresponding to the radius of a hole of cross-section area equal to the area of the annulus delimited by the pressure sensor and the circular edge of the hole. Specifically,

$$a_{\text{equivalent}} = \sqrt{a_{\text{hole}}^2 - a_{\text{sensor}}^2}, \tag{3}$$

with $a_{\text{sensor}}=145/2=72.5$ μm the radius of the pressure sensor.

During the experiments, making the hole usually resulted in perilymph leaking out from the cochlea at a slow rate. Therefore, the termination impedance $\mathbf{z_0}$ that we used was the mass of the fluid terminating the tube.[6]

The results obtained with this model share similarities with the experimental data for both $\mathbf{P_V}$ and $\mathbf{U_S}$ [Fig. 7(B)]. Introducing a 180 μm diameter hole (corresponding to the smallest hole size we achieved experimentally, and leading to $a_{\text{equivalent}}=53$ μm) reduced $|\mathbf{P_V}|$ near 150 Hz by about 10–12 dB, which is consistent with the 10–15 dB increase in the experimental data upon introduction of the gel to seal the hole. The effect of the hole was smaller as frequency increased, with less than a 3 dB difference by 1 kHz in both the experimental and predicted data. Nonetheless, the detailed shape of the predicted change in $|\mathbf{P_V}|$ is different from the measured change, where much of the differences come from frequency-dependent variations in $\mathbf{Z_C}$ and $\mathbf{Z_{out}}$ that originate in the details of the data used in their calculation (Songer and Rosowski, 2007a). As for the phase, the ~0.15 cycles increase predicted by the model at 150 Hz is consistent with the experimental data around this frequency, but the measured and predicted changes differ slightly at other frequencies.

The predicted changes in $\mathbf{U_S}$ are very small: The change in phase was close to 0 over the entire frequency range of the data (except the first data point at 62 Hz), and the change in magnitude was less than 1 dB below 1.5 kHz, and between 1 and 2 dB at frequencies 1.5–8 kHz (except for a small notch of −1 dB at 2.5 kHz). This is consistent with the experimen-

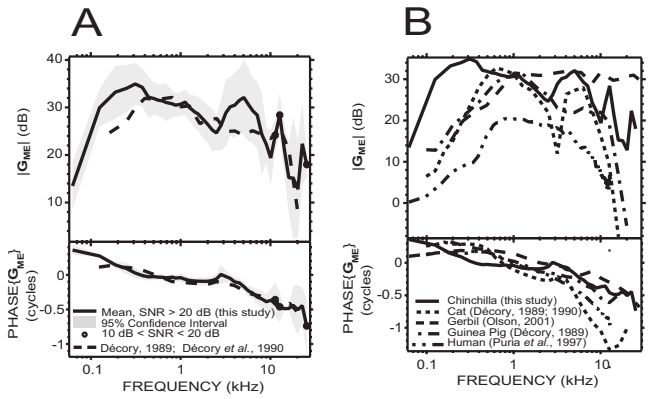


FIG. 8. Comparison of our measurements of $\mathbf{G_{ME}}$ (solid lines) with other studies. (A) In chinchilla: comparison with a previous study by Décory (1989) and Décory *et al.* (1990): dashed line. The two data sets were very similar at stimulus frequencies of 500 Hz–3 kHz and 12–20 kHz, but different at other frequencies. (B) In other mammals: comparison with studies in cat, guinea pig (from Décory, 1989 and Décory *et al.*, 1990: dotted and dashed-dotted lines), gerbil (from Olson, 2001: dashed line), and human temporal bone (from Puria *et al.*, 1997: dashed-double dotted line). All these measurements were carried out with open holes in the bullar walls. The septum of the middle ear (Møller, 1965; Huang *et al.*, 1997) was intact in the cat study.

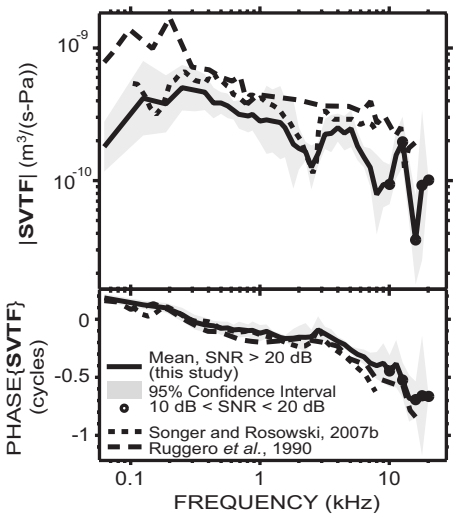


FIG. 9. Comparison of our measurements of **SVTF** (solid line) with previous studies in chinchilla by Songer and Rosowski (2007a): dotted line and Ruggero *et al.* (1990): dashed line. The differences in magnitude between ours and the Ruggero *et al.* (1990) study may be due to differences in the experimental setup (see text).

tal data over a wide range of frequencies: The average change in six animals was about 5 dB in magnitude below 6 kHz, but not statistically significant at these frequencies, and the phase was close to 0. Nonetheless, the slightly larger and significant experimental changes obtained around 6 kHz were not predicted.

To conclude, the predictions of this simple model are qualitatively similar to the experimental data in frequency dependence, and the predicted and measured changes in magnitude and phase angle are quantitatively similar: The model predicts that the small acoustic leak created by the vestibular hole would change $\mathbf{P_V}$ magnitude by no more than 12 dB at low frequencies, and phase angle by generally less than 0.15 cycles, and would cause only very small changes in $\mathbf{U_S}$ over 62 Hz–8 kHz.

E. Comparison with other studies in chinchilla

Our $\mathbf{G_{ME}}$ results are very similar to a previous study in chinchilla by Décory (1989; Décory *et al.*, 1990) between 500 Hz and 3 kHz, and between 8 kHz and 20 kHz, for both the magnitude and the phase [Fig. 8(A)]. Moreover, the slightly negative slope of $|\mathbf{G_{ME}}|$ between 500 Hz and 3 kHz, as well as the sharp decrease between 12 kHz and 20 kHz, was similar in both studies. Between 500 Hz and 3 kHz, the phase we measured was closer to 0 than Décory's (1989; Décory *et al.*, 1990), but the two did not differ by more than 0.1 cycles. Below 500 Hz, and between 3 and 8 kHz, we found a larger $|\mathbf{G_{ME}}|$, by as much as 7 dB at 250 Hz and 8 dB at 5 kHz. The low frequency difference may be explained by a difference in the acoustic leak introduced by the vestibular hole (see Sec. IV D). The difference near 5 kHz may be due to the animal preparation in Décory's (1989; Décory *et al.*, 1990) study: In their study, the vestibular hole was drilled about 5.5 mm away from the base of the cochlea, at a location corresponding to the 5 kHz place along the tonotopic axis. Therefore, as they propose, their measurements around 5 kHz may have been biased by a component due to the peak of the basilar membrane traveling wave, whereas our measurements were performed through a hole closer to the oval window (within 1–2 mm), far from the peak of the traveling wave over most of the measurement range (Rhode and Recio, 2000).

Our measurements of **SVTF** are similar to those in a study by Songer and Rosowski (2007a) in magnitude as well as phase (Fig. 9). Our results are also roughly similar to those of Ruggero *et al.* (1990) at frequencies below 12 kHz. However, we generally obtained a smaller $|\mathbf{SVTF}|$, especially below 300 Hz [$(2-5)\times10^{-10}$ $m^3/(s\ Pa)$ in our study, and 6×10^{-10} to 2×10^{-9} in Ruggero *et al.*, 1990] and near 3 and 8 kHz where we can see small notches in our study. These differences in magnitude may be due to the correction Ruggero *et al.* (1990) applied to take into account that the tensor tympani muscle was cut. Differences in the experimental setup may also explain some variations: We measured $\mathbf{V_S}$ with a laser Doppler vibrometer directed through the bulla, whereas Ruggero *et al.* (1990) used a Mössbauer source placed on the stapes footplate through a small slit in the TM. The bulla was widely open in both studies; however, the notch in our data at 2.5 kHz is likely the result of a bulla cavity-bullar-hole resonance (Rosowski *et al.*, 2006). The notch we found between 14 and 18 kHz is in contradiction with another study by Ruggero *et al.* (2007), in which they measured ossicular vibrations in chinchillas up to 40 kHz and obtained a roughly flat **SVTF** magnitude at least up to 25 kHz. As we discussed earlier, our high frequency results are not as reliable as the lower frequency range of our data, which could explain the difference. Another potential reason for these differences at high frequency is that Ruggero *et al.* (2007) measured velocity of the lenticular process, and added gains measured across the incudo-stapedial joint, whereas we measured velocity from locations on the footplate and parts of the crura close to the footplate.

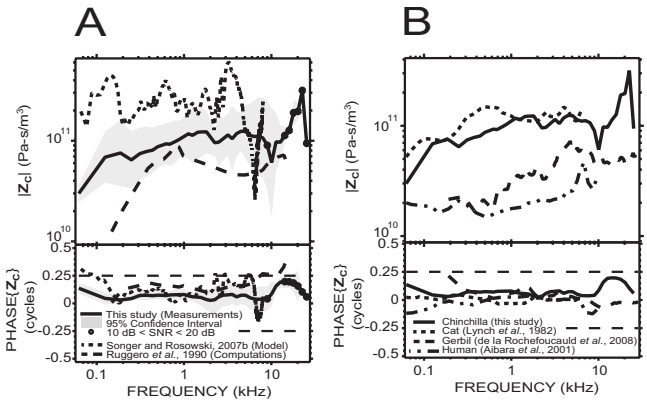


FIG. 10. Comparison of our measurements of $\mathbf{Z_C}$ (solid lines) with other studies. (A) In chinchilla: comparison with the prediction of a model by Songer and Rosowski (2007a): dotted line and computations by Ruggero *et al.* (1990): dashed line, who used their own $\mathbf{U_S}$ measurements and Décory's (1989) $\mathbf{P_V}$ measurements in other animals [see Fig. 8(A)]. (B) In other mammals: comparison with studies in cat (from Lynch *et al.*, 1982: dotted line), gerbil (from de La Rochefoucauld *et al.*, 2008: dashed line), and human temporal bone (from Aibara *et al.*, 2001: dashed-double dotted line).

We compared our measured $\mathbf{Z_C}$ with previous estimates in chinchilla by Ruggero *et al.* (1990) and Songer and Rosowski (2007a). Our $\mathbf{Z_C}$ is a more direct estimate than either of these previous studies because we derived $\mathbf{Z_C}$ from simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ whereas (1) Ruggero *et al.* (1990) combined their own $\mathbf{V_S}$ measurements (see Fig. 9) with Décory's (1989; Décory *et al.*, 1990) $\mathbf{P_V}$ measurements in other animals [see Fig. 8(A)], and (2) Songer and Rosowski (2007a) computed $\mathbf{Z_C}$ from a transmission matrix model of the middle ear fed by measurements of ear canal sound pressure and stapes velocity. The three data sets share many similarities [Fig. 10(A)]: In particular, the impedances are mostly resistive (phase angles close to 0 and approximately flat magnitude on a broad frequency range) with a magnitude of about 10^{11} acoustic ohms; the phase angles of all three estimates are similar at frequencies between 0.3 and 4 kHz. There are also marked differences.

(1) Ruggero *et al.*'s (1990) estimate has a lower magnitude than ours below 500 Hz and between 3 and 10 kHz, consistent with the larger $|\mathbf{G_{ME}}|$ we measured at these frequencies. Their phase angles are higher, and take values greater than +0.25 cycles above 7 kHz, whereas our estimate remains within ±0.25 cycles at all frequencies.
(2) There is a consistent factor of 2–3 difference between the impedance computed by Songer and Rosowski (2007a) and our direct measurements. Their phase angle, although generally near 0, displays some irregularities near 3 kHz and near 7 kHz.

F. Comparison with other mammals

$\mathbf{G_{ME}}$ is shown for chinchilla (our data) along with cat and guinea pig (from Décory, 1989; Décory *et al.*, 1990), gerbil (from Olson, 2001), and human temporal bone (from Puria *et al.*, 1997) in Fig. 8(B). All these data were measured with the middle ear space open. The septum was intact in the cat study. $|\mathbf{G_{ME}}|$ is largest for the chinchilla, especially at low frequencies, but $|\mathbf{G_{ME}}|$ in all these species is similar (within 10 dB) for frequencies between 500 Hz and 3 kHz.

Except for human temporal bone, the overall shape of $|\mathbf{G_{ME}}|$ can be roughly described in these species by two more-or-less broad lobes. For chinchilla, the first lobe ranges from 62 Hz to 2.5 kHz and peaks at 35 dB, while the second one ranges from 2.5 to 9 kHz and peaks at a slightly lower value (32 dB); at higher frequencies, the gross tendency is a decrease in gain, but there is no evidence of a sharp roll-off. For cat and guinea pig, the first lobe has a larger maximum than the second one (about 32 dB for cat and 31 dB for guinea pig). The separation between the two lobes is more prominent in the cat data (the large notch centered at 3 kHz is related to a resonance between the tympanic middle ear cavity and the foramen in the septal wall of the bulla, e.g., Møller, 1965; Guinan and Peake, 1967; Huang *et al.*, 1997). There are sharp high frequency roll-offs for cat and guinea pig. For the cat, the roll-off occurs at slightly lower frequencies (0 dB is reached by 15 kHz) than for guinea pig. For the gerbil data, the separation between the two lobes is at about 7 kHz, but the second lobe extends to at least 46 kHz [data not shown in Fig. 8(B)] and there is no evidence of a roll-off at these frequencies. The more-or-less pronounced notch separating the two lobes in the animal data described above is likely due to a resonance between the bulla and the open hole in the bullar wall (e.g., Møller, 1965; Ravicz *et al.*, 1992; Rosowski *et al.*, 2006).

The phases of $\mathbf{G_{ME}}$ in these species have similarities in shape, but the decrease with frequency varies across species (fastest for the cat and slowest for the gerbil). The average group delay between 0.2 and 10 kHz is about 52 μs in our study, 131 μs in cat, 51 μs in gerbil, 82 μs in guinea pig, and 92 μs in human. For cat and guinea pig, the phase increases slightly at high frequency after reaching a minimum. It is difficult to tell whether this increase is real or if it is an artifact of phase unwrapping.

We compared our $\mathbf{Z_C}$ results with other studies where $\mathbf{P_V}$ and $\mathbf{V_S}$ were simultaneously measured [Fig. 10(B)] in cat (from Lynch *et al.*, 1982), gerbil (from de La Rochefoucauld *et al.*, 2008), and human temporal bone (from Aibara *et al.*, 2001). $\mathbf{Z_C}$ in chinchilla and cat are very similar up to 8 kHz for both magnitude and phase. For all these species, the magnitudes are approximately flat with frequency and the phase angles close to 0. The mostly resistive cochlear input impedance in all those mammalian species ensures that most of the energy transmitted through the middle ear is dissipated into the cochlear fluid, and may be important for cochlear traveling wave propagation (Décory, 1989; de La Rochefoucauld *et al.*, 2008).

G. Acoustic power W_C, $\mathbf{P_V}$, $\mathbf{U_S}$, and the audiogram

We tested the hypothesis that the shape of the audiogram is set by a constant threshold value of some input quantity such as the average acoustic power delivered to the cochlea W_C (Khanna and Tonndorf, 1969; Rosowski, 1991), $\mathbf{P_V}$, or $\mathbf{U_S}$.

W_C can be computed from $\mathbf{P_V}$ and the cochlear input impedance $\mathbf{Z_C}$ as follows:

$$W_C = \frac{1}{2}\mathrm{Re}\{\mathbf{P_V}\mathbf{U_S}^*\} = \frac{1}{2}|\mathbf{P_V}|^2\mathrm{Re}\left\{\frac{1}{\mathbf{Z_C}}\right\}, \quad (4)$$

where Re{**X**} indicates the real part of **X**. In terms of our measured quantities,

$$\frac{W_C}{|\mathbf{P_{TM}}|^2} = \frac{1}{2}\left|\frac{\mathbf{P_V}}{\mathbf{P_{TM}}}\right|^2\mathrm{Re}\left\{\frac{1}{\mathbf{Z_C}}\right\} = \frac{1}{2}|\mathbf{G_{ME}}|^2\mathrm{Re}\left\{\frac{1}{\mathbf{Z_C}}\right\}, \quad (5)$$

and the sound pressure at the TM $\mathbf{P_{TM,0}}$ required to produce a criterion value of acoustic power $W_{C,0}$ is

$$|\mathbf{P_{TM,0}}| = \sqrt{\frac{W_{C,0}}{\frac{1}{2}|\mathbf{G_{ME}}|^2\mathrm{Re}\left\{\frac{1}{\mathbf{Z_C}}\right\}}}. \quad (6)$$

By definition of $\mathbf{G_{ME}}$ and **SVTF**, the sound pressure at the TM $\mathbf{P_{TM,0}}$ required to produce a criterion value of vestibular pressure $\mathbf{P_{V,0}}$ or of stapes volume velocity $\mathbf{U_{S,0}}$ is given by

$$\mathbf{P_{TM,0}} = \frac{\mathbf{P_{V,0}}}{\mathbf{G_{ME}}}, \quad (7)$$

$$\mathbf{P_{TM,0}} = \frac{\mathbf{U_{S,0}}}{\mathbf{SVTF}} \quad (8)$$

We compare free-field sound pressure at auditory threshold in chinchilla (Miller, 1970) with the sound pressure required to produce a constant level of W_C [by Eq. (6)], $\mathbf{P_V}$ [by Eq. (7)], and $\mathbf{U_S}$ [by Eq. (8)]. A complication is that the chinchilla audiogram from Miller (1970) was measured in free field, whereas our experiments were done with the sound stimuli specified at the TM. To account for the differences between free-field pressure $\mathbf{P_{FF}}$ and ear canal pressure near the TM $\mathbf{P_{TM}}$, we used an average head related transfer function (HRTF) measured in the chinchilla by von Bismark and Pfeiffer (1967). This HRTF quantifies the filtering effect of the head and pinnae on the incoming sound. Another complication is that all our measurements were done with the bulla open, whereas Miller's (1970) audiogram was measured in behaving animals with an intact bulla. We used a correction factor from Ruggero *et al.* (1990) to account for the effect of opening the bulla on $\mathbf{G_{ME}}$ and **SVTF**: $|\mathbf{G_{ME}}|$ and $|\mathbf{SVTF}|$ in the open bulla condition were larger than in the closed bulla condition by as much as 16 dB in the low frequencies; the effect was less than 3 dB above 600 Hz.

Figure 11 compares three free-field sound pressure estimates $|\mathbf{P_{FF}}|$ to Miller's (1970) audiogram: the $|\mathbf{P_{FF}}|$ required to generate $W_{C,0}=5\times10^{-18}$ W of acoustic power in the vestibule, the $|\mathbf{P_{FF}}|$ required to generate $|\mathbf{P_{V,0}}|=34$ dB re 20 μPa, and the $|\mathbf{P_{FF}}|$ required to generate $|\mathbf{U_{S,0}}|=10^{-14}$ m^3/s. These values were chosen by eye for a close fit to the audiogram over the frequency range 125 Hz–2 kHz. The HRTF was available only between 250 Hz and 8 kHz, which limited the range of $\mathbf{P_{FF}}$. Nonetheless, given that the wavelength of sound at low frequencies (>1.4 m below 250 Hz in air) is large compared to the size of the head and pinnae, we can assume that the HRTF has a 0 dB gain below 250 Hz, and therefore $\mathbf{P_{TM}}$ well approximates $\mathbf{P_{FF}}$ below 250 Hz.

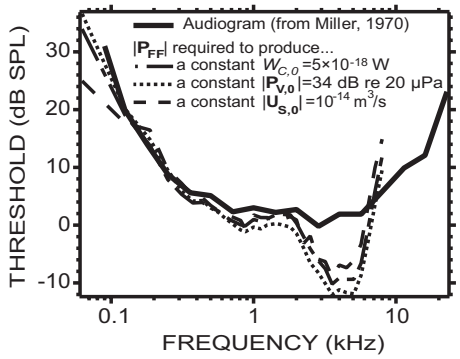


FIG. 11. An average chinchilla audiogram (thick solid line, from Miller, 1970) was compared to the free-field sound pressure $|\mathbf{P_{FF}}|$ required to produce a constant level of $|\mathbf{P_V}|$ (thin dotted line), $|\mathbf{U_S}|$ (thin dashed line), and W_C (thin dashed-dotted line).

The audiogram is well predicted by the three quantities tested between 125 Hz and 2 kHz. Below 125 Hz, both W_C and $\mathbf{P_V}$ produce similar slopes to the audiogram, whereas $\mathbf{U_S}$ produces a shallower slope and therefore is not a good predictor at those frequencies. Above 2 kHz, none of the three quantities are good predictors: The audiogram is almost flat with thresholds between 0 and 5 dB SPL, whereas $\mathbf{P_{FF}}$ in the three cases decreases to a minimum near 2 kHz before increasing again with frequency, crossing the audiogram at about 7 kHz.

In Rosowski (1991), W_C was estimated from a power flow model using various mechano-acoustic quantities in cat, chinchilla, and human temporal bone. The $\mathbf{P_{FF}}$ required to produce a constant $W_C=5\times10^{-18}$ W was plotted against Miller's (1970) audiogram in the case of the chinchilla (see his Fig. 12, middle panel). The fit between $\mathbf{P_{FF}}$ and the audiogram was determined so as to minimize the mean square error over the full range of tested frequencies, whereas we minimized (by eye) the error below 2 kHz (which by chance led to the same value of 5×10^{-18} W). In both studies, W_C is a poor predictor of the audiogram above 2 kHz. Below 2 kHz, the shape of our $\mathbf{P_{FF}}$ more reliably predicts the shape of the audiogram.

V. CONCLUSIONS

We used the miniature fiber-optic pressure sensors described by Olson (1998) to measure the sound pressure within the vestibule $\mathbf{P_V}$ of the inner ears of chinchillas, and laser Doppler vibrometry to measure stapes velocity $\mathbf{V_S}$, while presenting sound stimuli to the ear canal (Fig. 1). Our simultaneous measurements of $\mathbf{P_V}$ and $\mathbf{V_S}$ led to new estimates of the middle ear pressure gain $\mathbf{G_{ME}}$ and of the stapes velocity transfer function **SVTF** in chinchillas, and to the first direct measurement of cochlear input impedance $\mathbf{Z_C}$ in this species. Measurements of $\mathbf{V_S}$ before and after introducing the hole in the bony vestibular wall and placing the fiber-optic pressure sensor in the vestibule (Fig. 4), as well as measurements of $\mathbf{P_V}$ before and after sealing the pressure sensor in place (Fig. 3), suggest that the introduction of the vestibular hole and placement of the sensor had a small effect on middle ear sound transmission. We were able to

qualitatively predict this effect using a simple acoustic model (Fig. 7).

The measured $\mathbf{G_{ME}}$ (Fig. 2) was band-pass in nature with a peak magnitude similar to the 38 dB middle ear gain predicted from the product of the TM/footplate area ratio and the malleus/incus lever ratio. The phase angle of $\mathbf{G_{ME}}$ varied by more than a cycle over the 62 Hz–30 kHz measurement range and was generally consistent with a group delay of 52 μs. When compared with cat, gerbil, guinea pig, and human temporal bone (Fig. 8), $|\mathbf{G_{ME}}|$ was largest for the chinchilla, especially at low frequencies. However, $|\mathbf{G_{ME}}|$ in all these species was similar (within 10 dB) over 500 Hz–3 kHz.

Our $\mathbf{Z_C}$ was resistive from 0.2 to at least 10 kHz, with a magnitude on the order of 10^{11} acoustic ohms (Fig. 5). Together with previous direct measurements of $\mathbf{Z_C}$ in cat, gerbil, and human temporal bone (Fig. 10), this suggests that a resistive $\mathbf{Z_C}$ is characteristic of the mammalian ear. Over 125 Hz–2 kHz, estimates of the free-field sound pressures necessary to maintain a constant sound power into the inner ear, a constant $\mathbf{P_V}$, or a constant $\mathbf{V_S}$ had a frequency dependence similar to that of auditory thresholds (Fig. 11); this similarity did not hold at higher sound frequencies. A better predictor of the shape of the audiogram may be the basilar membrane differential pressure (difference between $\mathbf{P_V}$ and the sound pressure in scala tympani), which will be the object of a future study in a similar chinchilla preparation.

ACKNOWLEDGMENTS

We thank Elizabeth Olson, Wei Dong, and Heidi Nakajima, for their assistance in making the fiber-optic pressure sensors and their advice in using them in our experiments. We also thank Melissa Wood for her mastery of chinchilla surgery and her help with the experiments. We would also like to express our appreciation to the staff of the Microsystems Technology Laboratories at MIT and, in particular, to Kurt Broderick. This work was supported by Grant Nos. RO1-DC00194 and T32-DC00038 from the National Institute on Deafness and other Communication Disorders (NIDCD).

[1]In this paper, bold variables ($\mathbf{X}$) are complex functions of frequency and variables in italics (Y) are real.

[2]The average group delays reported in this paper over an x Hz$-y$ Hz frequency range were computed as $-(\phi(y)-\phi(x))/(y-x)$, where $\phi(z)$ is the phase angle in cycle at frequency z Hz.

[3]The theoretical anatomical "transformer ratio" was computed as the product of the "area ratio" (the area of the TM divided by the area of the stapes footplate) and the "lever ratio" (malleus length divided by incus length). Anatomical values in the chinchilla lead to an "area ratio" of 29 dB (using an estimated TM area of about 60 mm^2 and a footplate area of 2 mm^2 from Vrettakos *et al.*, 1988) while differences in malleus and incus length suggest a "lever ratio" of 9 dB (using 4.5 mm as the estimated length of the manubrium of the malleus and 1.58 mm as the estimated length of the long process of the incus, from Vrettakos *et al.*, 1988). The total "transformer ratio" is therefore 38 dB.

[4]We did not use our own $\mathbf{Z_C}$ because it was measured with a hole in the vestibule.

[5]We also tried a simpler model of $\mathbf{Z_{hole}}$ to predict the change in $\mathbf{P_V}$ and $\mathbf{U_S}$: We used equations for a tube of very small diameter from Beranek (1986) [Eqs. (5.48) and (5.49), p. 135]. Even though this simpler model has some frequency limitations, the predictions were very similar to those based on the lossy transmission line model of $\mathbf{Z_{hole}}$ from Eq. (2).

[6]We used the termination impedance for a tube terminating in an infinite baffle from Beranek (1986) [Eq. (5.17), p. 121]. The acoustic mass obtained for a 180 μm diameter hole, leading to an equivalent radius $a_{\text{equivalent}}=53$ μm, was about 5.1×10^6 kg/m^4.

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**, 100–109.

Beranek, L. L. (**1986**). *Acoustics* (Acoustical Society of America, Melville, NY).

Chien, W., Ravicz, M. E., Merchant, S. N., and Rosowski, J. J. (**2006**). "The effect of methodological differences in the measurement of stapes motion in live and cadaver ears," Audiol. Neuro-Otol. **11**, 183–197.

Dallos, P. (**1973**). *The Auditory Periphery* (Academic, New York).

Dancer, A., and Franke, R. (**1980**). "Intracochlear sound pressure measurements in guinea pigs," Hear. Res. **2**, 191–205.

de La Rochefoucauld, O., Decraemer, W. F., Khanna, S. M., and Olson, E. S. (**2008**). "Simultaneous measurements of stapes motion and intracochlear pressure in gerbil from 0.5–50 kHz," J. Assoc. Res. Otolaryngol. **9**, 161–77.

Décory, L. (**1989**). "Origine des différences interspécifiques de susceptibilités au bruit (Origins of interspecies differences in susceptibility to noise)," Ph.D. thesis, Université de Bordeaux, France.

Décory, L., Franke, R. B., and Dancer, A. L. (**1990**). in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer, Berlin), pp. 270–277.

Decraemer, W. F., de La Rochefoucauld, O., Dong, W., Khanna, S. M., Dirckx, J. J., and Olson, E. S. (**2007**). "Scala vestibuli pressure and three-dimensional stapes velocity measured in direct succession in gerbil," J. Acoust. Soc. Am. **121**, 2774–2791.

Decraemer, W. F., Dirckx, J. J. J., and Funnell, W. R. J. (**2003**). "Three-dimensional modelling of the middle ear ossicular chain using a commercial high-resolution x-ray CT scanner," J. Assoc. Res. Otolaryngol. **4**, 250–263.

Decraemer, W. F., and Khanna, S. M. (**2003**). "Measurement, visualization and quantitative analysis of complete three-dimensional kinematical data sets of human and cat middle ear," Proceedings of the Middle Ear Mechanics in Research and Otology, Matsuyama, Japan, pp. 3–10.

Dong, W., and Olson, E. S. (**2006**). "Middle ear forward and reverse transfer function," J. Neurophysiol. **95**, 2951–2961.

Egolf, D. P. (**1977**). "Mathematical modeling of a probe-tube microphone," J. Acoust. Soc. Am. **61**, 200–205.

Funnell, W. R. J. (**1996**). "Low-frequency coupling between eardrum and manubrium in a finite-element model," J. Acoust. Soc. Am. **99**, 3036–3043.

Guinan, J. J., and Peake, W. T. (**1967**). "Middle-ear characteristics of anesthetized cats," J. Acoust. Soc. Am. **41**, 1237–1261.

Hato, N., Stenfelt, S., and Goode, R. L. (**2003**). "Three-dimensional stapes footplate motion in human temporal bones," Audiol. Neuro-Otol. **8**, 140–52.

Heiland, K. E., Goode, R. L., Asai, M., and Huber, A. M. (**1999**). "A human temporal bone study of stapes footplate movement," Am. J. Otol. **20**, 81–6.

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch, T. J., III, and Peake, W. T. (**1997**). "The middle ear of a lion: Comparison of structure and function to domestic cat," J. Acoust. Soc. Am. **101**, 1532–1549.

Khanna, S. M., and Tonndorf, J. (**1969**). "Middle ear power transfer," Arch. Klin. Exp. Ohren Nasen Kehlkopfheilkd **193**, 78–88.

Lynch, T. J., III, Nedzelnitsky, V., and Peake, W. T. (**1982**). "Input impedance of the cochlea in cat," J. Acoust. Soc. Am. **72**, 108–130.

Magnan, P., Avan, P., Dancer, A., Smurzynski, J., and Probst, R. (**1997**). "Reverse middle ear transfer function in the guinea pig measured with cubic difference tones," Hear. Res. **107**, 41–45.

Miller, J. D. (**1970**). "Audibility curve of the chinchilla," J. Acoust. Soc. Am. **48**, 513–523.

Møller, A. R. (**1965**). "Experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Oto-Laryngol. **60**, 129–149.

Nakajima, H. H., Dong, W., Olson, E. S., Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (**2009**). "Differential intracochlear sound pressure measurements in normal human temporal bones," J. Assoc. Res. Otolaryngol. **10**, 23–36.

Nedzelnitsky, V. (**1980**). "Sound pressures in the basal turn of the cat cochlea," J. Acoust. Soc. Am. **68**, 1676–1689.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel

intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.
Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.
Overstreet, E. H., and Ruggero, M. A. (**2002**). "Development of wide-band middle ear transmission in the Mongolian gerbil," J. Acoust. Soc. Am. **111**, 261–270.
Peake, W. T., Rosowski, J. J., and Lynch, T. J., III (**1992**). "Middle-ear transmission: Acoustic vs. ossicular coupling in cat and human," Hear. Res. **57**, 245–268.
Plontke, S. K. R., Wood, A. W., and Salt, A. N. (**2002**). "Analysis of gentamicin kinetics in fluids of the inner ear with round window administration," Otol. Neurotol. **23**, 967–74.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Puria, S., and Allen, J. B. (**1998**). "Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay," J. Acoust. Soc. Am. **104**, 3463–3481.
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.
Ravicz, M. E., Cooper, N. P., and Rosowski, J. J. (**2008**). "Gerbil middle ear sound transmission from 100 Hz to 60 kHz," J. Acoust. Soc. Am. **124**, 363–380.
Ravicz, M. E., Rosowski, J. J., and Voigt, H. F. (**1992**). "Sound-power collection by the auditory periphery of the Mongolian gerbil *Meriones unguiculatus*. I: Middle ear input impedance," J. Acoust. Soc. Am. **92**, 157–177.
Ravicz, M. E., Slama, M. C. C., and Rosowski, J. J. (**2009**). "Middle-ear pressure gain and cochlear partition differential pressure in chinchilla," Hearing Res. (in press).
Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.
Rosowski, J. J. (**1991**). "The effects of external and middle ear filtering on auditory threshold and noise-induced hearing loss," J. Acoust. Soc. Am. **90**, 124–135.
Rosowski, J. J. (**1994**). in *Springer Handbook of Auditory Research: Comparative Hearing: Mammals*, edited by A. Popper and R. Fay (Springer-Verlag, New York), Vol. **IV**, pp. 172–247.
Rosowski, J. J., Ravicz, M. E., and Songer, J. E. (**2006**). "Structures that contribute to middle ear admittance in chinchilla," J. Comp. Physiol. [A] **192**, 1287–1311.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). "Middle ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," J. Acoust. Soc. Am. **87**, 1612–1629.
Ruggero, M. A., Temchin, A. N., Fan, Y.-H., and Cai, H. (**2007**). "Boost of transmission at the pedicle of the incus in the chinchilla middle ear," 4th International Symposium on Middle Ear Mechanics in Research and Otology, pp. 154–157.
Schloss, F., and Strasberg, M. (**1962**). "Hydrophone calibration in a vibrating column of liquid," J. Acoust. Soc. Am. **34**, 958–960.
Shera, C. A., and Zweig, G. (**1992**). "Middle ear phenomenology: The view from the three windows," J. Acoust. Soc. Am. **92**, 1356–1370.
Songer, J. E., and Rosowski, J. J. (**2005**). "The effect of superior canal dehiscence on cochlear potential in response to air-conducted stimuli in chinchilla," Hear. Res. **210**, 53–62.
Songer, J. E., and Rosowski, J. J. (**2006**). "The effect of superior-canal opening on middle ear input admittance and air-conducted stapes velocity in chinchilla," J. Acoust. Soc. Am. **120**, 258–269.
Songer, J. E., and Rosowski, J. J. (**2007a**). "Transmission matrix analysis of the chinchilla middle ear," J. Acoust. Soc. Am. **122**, 932–942.
Songer, J. E., and Rosowski, J. J. (**2007b**). "A mechano-acoustic model of the effect of superior canal dehiscence on hearing in chinchilla," J. Acoust. Soc. Am. **122**, 943–951.
von Bismark, G., and Pfeiffer, R. R. (**1967**). "On the sound pressure transformation from free field to eardrum of chinchilla," J. Acoust. Soc. Am. **42**, S156.
von Helmholtz, H. L. F. (**1877**). *On the Sensations of Tones as a Physiological Basis for the Theory of Music* (Dover, New York).
Vrettakos, P. A., Dear, S. P., and Saunders, J. C. (**1988**). "Middle ear structure in the chinchilla: A quantitative study," Am. J. Otolaryngol. **9**, 58–67.
Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton University Press, Princeton, NJ).

The effect of tectorial membrane and basilar membrane longitudinal coupling in cochlear mechanics

Julien Meaud[a)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109

Karl Grosh
Department of Mechanical Engineering and Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

(Received 25 August 2009; revised 15 December 2009; accepted 15 December 2009)

Most mathematical models of the mammalian cochlea neglect structural longitudinal coupling. However, recent experimental data suggest that viscoelastic longitudinal coupling, in the basilar membrane (BM) and the tectorial membrane (TM), is non-negligible. In this paper, mathematical models for BM and TM longitudinal coupling are presented to determine the influence of such a coupling on the tuning of the BM. The longitudinal coupling models are added to a macroscopic linear model of the guinea pig cochlea that includes the micromechanics of the organ of Corti and outer hair cell (OHC) somatic motility. The predictions of the BM response to acoustic stimulus show that the characteristic frequency is controlled by a TM radial resonance and that TM longitudinal coupling has a more significant effect than BM longitudinal coupling. TM viscoelasticity controls the sharpness of the BM frequency response and the duration of the impulse response. The results with realistic TM longitudinal coupling are more consistent with experiments. The model predicts that OHC somatic electromotility is able to supply power to the BM at frequencies well above the cutoff of the OHC basolateral membrane. Moreover, TM longitudinal coupling is predicted to stabilize the cochlea and enable a higher BM sensitivity to acoustic stimulation. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290995]

PACS number(s): 43.64.Kc, 43.64.Bt [BLM] Pages: 1411–1421

I. INTRODUCTION

The mammalian hearing system combines high sensitivity to low level acoustic pressure stimulus with a dynamic range that extends over six orders of magnitude. In addition, cochlear responses are highly tuned in the frequency domain yet the system as a whole still possesses excellent transient capture, able to discriminate timing differences of $6-10\ \mu s$.[1] The solution to these seemingly conflicting characteristics involves both the unique transduction properties of the auditory periphery and the processing capabilities of the central nervous auditory system. In the periphery, where the acoustic signals are converted to neuronal input, an intricate micromechanical and microfluidic cochlear anatomy has evolved. In the healthy cochlea, outer hair cells (OHCs) present a nonlinear electrical and mechanical response to acoustic stimulation. The mechanical force from the OHC is thought to be the main factor leading to both the nonlinear input-output characteristic and the sharp frequency filtering seen in the cochlea. The focus of current research aimed at uncovering the workings of the cochlea has been on two mechanisms of OHC mediated force generation, basolateral (somatic) and hair bundle (HB) motility. Both hypotheses hinge on the conversion of some form of stored nonmechanical energy (e.g., the endocochlear electrical potential[2]) to mechanical energy. The HBs are comprised of numerous stereocilia-like projections from the apex of each OHC. The HBs are of central interest, as shear deflection of the HB gates the large potassium current necessary for somatic OHC force generation, and the same shear motion of the HB is thought to initiate a cascade of events resulting in HB force generation. The apical termination of the tallest row of stereocilia of each HB is in the tectorial membrane (TM) which, therefore, plays a critical role in active cochlear mechanics. The basilar membrane (BM) is a main structural component of the cochlea since it is directly coupled to the fluid of the cochlear ducts. The sensory hair cells are sandwiched between the TM and the BM. In this paper we use a mathematical model of the cochlea that explicitly includes the micromechanics of the organ of Corti (OoC) with independent degrees of freedom for the BM and TM vibrations. We introduce longitudinal coupling in the TM and/or in the BM using material properties based on experimental data to predict the effect of such coupling on the BM response to acoustic input.

Some researchers have postulated a central role for the TM in cochlear mechanics, and some before the electromotility of the OHC had been presented in 1985.[3] Zwislocki[4,5] hypothesized that the TM acts as a second resonator coupled to the BM through the OHC HB linkage. He used this to explain the sharp tuning and secondary peaks in tuning curve. Allen[6] postulated that a two degree of freedom resonator system consisting of the TM mass and BM-OoC mass would be sufficient to explain the sharpness seen in cochlear tuning.

a)Author to whom correspondence should be addressed. Electronic mail: jmeaud@umich.edu

Gummer *et al.*[7] presented experimental results on TM resonance in the apex of the postmortem cochlea. If these results can be extended to the basal region of the living cochlea, they indicate the presence of two TM modes having different resonant frequencies. Mammano and Nobili[8] developed a model of TM interaction with HB giving rise to OHC somatic electromotility. Chadwick *et al.*[9] also developed a model that highlighted the importance of the TM for predicting the frequency response of the cochlea. In the former models, certain assumptions are made regarding the amplitude and/or phase of the OHC somatic forcing that limit their predictive capability, especially with regard to the TM mechanics and OHC electromotility.

The TM is a gelatinous structure with three different noncollageneous glycoproteins (α-tectorin, β-tectorin, and otogelin). β-tectorin is an essential structural component providing longitudinal coupling in the TM (as shown by Ghaffari[10]). A mouse with genetically modified β-tectorin exhibits an enhanced tuning and reduced sensitivity in the high frequency region,[11] which suggests that the TM plays a key role in tuning and that TM longitudinal coupling is important for cochlear mechanics. The mechanical properties of the TM have been measured in gerbils[12,13] and in guinea pigs[14,15] as well as in mice.[16–18] The measurements in Ref. 17 show that the TM has frequency dependent properties. Moreover, although direct comparison between different measurements is difficult because of the difference in the experimental methods, the data show that the TM of mice (that have a higher frequency range than guinea pigs) is much stiffer than the TM of the guinea pigs and gerbils (that have a lower frequency range than guinea pigs). Zwislocki and Cefaratti[13] found that the TM is significantly less stiff than the HBs in the gerbil cochlea. However, some more recent measurements lead to contradictory conclusions about the relative stiffness of the TM (which in turn influences the kinematics). Some studies in the mouse cochlea have found that the TM is significantly stiffer than the HB.[16,17] If this is the case, then the TM would then move as rigid body from the limbal attachment provided that the TM region near the spiral limbus is more compliant than the main body of the TM (see Ref. 19). Recent measurements in the gerbil and guinea pig cochlea found that the TM has a stiffness within an order of magnitude of the HB.[12,15] In this case, the TM would then deform elastically from the limbal attachment (if the limbal attachment is stiffer than the TM, as suggested in Refs. 14 and 18). Richter *et al.*[12] showed that the TM stiffness varies longitudinally with a higher radial stiffness at basal locations than at more apical locations. Shoelson *et al.*[15] did not observe the presence of a stiffness gradient in the TM of the guinea pig but observed a longitudinal and radial inhomogeneity. An estimate of the shear modulus and Young's modulus was derived in Refs. 12 and 15 based on an isotropic model of the TM. As noted in these papers some caution must be taken in interpreting these results because of the isotropic and homogeneous assumptions of the model. Because of its microstructure the TM is anisotropic, as shown by Gavara and Chadwick[14] using atomic force microscopy measurements of the elastic moduli. Ghaffari *et al.*[20] demonstrated that the TM isolated from the mouse cochlea is capable of supporting shear waves that propagate in the longitudinal direction of the cochlea (see Fig. 1). Such waves would not be possible without significant longitudinal stiffness in the TM. While the mechanical properties of the TM estimated from these experiments are somewhat different with those from Gavara and Chadwick,[14] both group's estimates of modulus are consistent with a slow shear wave whose phase velocity is comparable to that of the traveling wave in the cochlea near the frequency dependent peak response location.

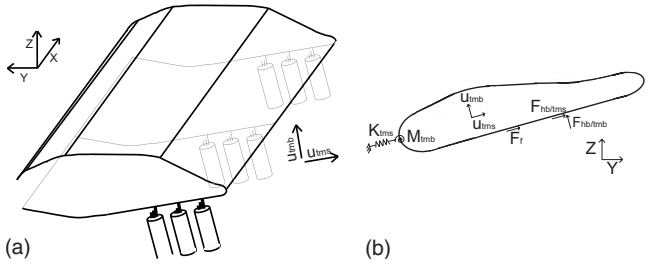


FIG. 1. (a) Illustration of the TM model. The x-coordinate corresponds to the longitudinal direction. (b) Cross-sectional view of the TM showing external forces (per unit length) acting on the TM. The HBs apply forces in the shear ($F_{hb/tms}$) and bending ($F_{hb/tmb}$) directions. Fluid forcing due to the viscous fluid interaction in the subtectorial space is given by a force ($F_f = C^f_{sub}\dot{u}_s$) in the shear direction proportional to the relative shear velocity between the TM and the RL, $\dot{u}_s$. The TM elasticity in the cross section is modeled by a linear spring (K_{tms}) applying a force in the shear direction and a rotational spring (M_{tmb}) applying a force in the bending direction. The plane sections (y-z) shear relative to one another. As in Ref. 20, internal viscous ($A^{eff}_{tm}\eta_{xy}\partial\dot{u}_{tms}/\partial x$) and elastic ($A^{eff}_{tm}G_{xy}\partial u_{tms}/\partial x$) coupling is included.

In the BM stiffness measurements in dead animals from von Bekesy,[21] longitudinal coupling appears to be significant. However, in the study from Voldrich[22] in live guinea pigs, longitudinal coupling in the BM is negligible. More recent work in mongolian gerbil cochlea[23] quantifies longitudinal coupling in the BM and organ of Corti and indicates that longitudinal coupling is significant and increases from the base to the apex of the cochlea. Liu and White[24] used the published experimental data to compute the material properties of the BM described by an orthotropic plate model.

Recent experimental data indicate that viscoelastic coupling in the TM and elastic coupling in the BM is important. However, most cochlear models[25,19] are based on a locally reacting representation of the cochlear partition and neglect structural longitudinal coupling. Allen and Sondhi[26] modeled the BM as an orthotropic plate but the TM is not included and the model is passive. The model of Wickersberg and Geisler[27] includes elastic longitudinal coupling in the BM in a one dimensional model. In this model the effect of introducing longitudinal coupling in the BM governing equation for the low damping case is to broaden the BM frequency response and to reduce the magnitude of the peak. But the interpretation of this result is limited since the organ of Corti is reduced to one degree of freedom and the magnitude of longitudinal coupling is not based on any measurement. Steele and Taber[28] included longitudinal coupling in the BM (treating it as a plate). The model of Mammano and Nobili[8] introduces only viscous longitudinal coupling in the organ of Corti. If structural longitudinal coupling is important in the

TABLE I. Tectorial membrane properties (x is in m).

Properties	Value	Ref.
A_{tm}^{eff}	$A_{tm}^{0}e^{50x}$ kg/m, $A_{tm}^{0}=2600$ μm^2	Based on Ref. 12
K_{tms}	$1.4\times10^{4}e^{-\alpha_{tm}x}$ N/m^2 with $\alpha_{tm}=300$ m^{-1}	Based on Ref. 12, see Appendix
M_{tms}	$0.9\rho_{tm}A_{tm}^{eff}$, $\rho_{tm}=1200$ kg/m^3	
G	$7.0e^{-\alpha_{tm}x}$ kPa	0.67 kPa at basal location and 0.44 kPa at an apical location in the guinea pig[a]
η	0.05 Pa s	0.20 Pa s in the mouse TM[b]

[a]Reference 14.
[b]Reference 20.

mechanics of the cochlea, what is the role of such coupling? We propose a role for the TM in electromotile processes and in shaping the frequency response of the BM to acoustic stimulus. We also predict the relative importance of the BM versus TM longitudinal coupling in influencing the sharpness of tuning of the BM.

II. METHODS

The fluid domain is idealized as uncoiled with two internal ducts separated by the basilar membrane. The macroscopic response of the fluid is modeled as inviscid and incompressible and is coupled to the mechanical response of the basilar membrane through the linearized Euler relation.[19] A micromechanical model of the organ of Corti is coupled mechanically to the TM and the BM. Viscous fluid effects (e.g., shearing of the fluid in the subtectorial space) are also included in the micromechanical model. Electrical conduction through the scalae is modeled using a set of coupled cable equations. The electrical and mechanical domains are coupled through a piezoelectric model for the OHC somatic motility and a displacement dependent conductance of the HB. Details of the mechanical, electrical, and fluidic model are presented by Ramamoorthy *et al.*[19] The main change to the model of Ramamoorthy *et al.* is the addition of longitudinal coupling in the TM and BM mechanics as is described next.

A. Tectorial membrane mathematical model

As shown in Fig. 1, each cross section of the TM is modeled as a two degree of freedom system with deformation in the radial or shear direction (u_{tms}) and in the normal or bending direction (u_{tmb}). The TM is characterized at each (y-z) plane by its effective stiffness and mass per unit length. The bending and shear motion varies in the longitudinal direction (x) and hence the cross sections move relative to one another. Longitudinal viscoelastic coupling of the shear motion of the TM is included (with a shear modulus G_{xy} and a shear viscosity η_{xy}) while TM bending rigidity is neglected. The governing equation for the shear motion of the TM is

$$F_{hb/tms}(x) = K_{tms}u_{tms} + C_{sub}^{f}\dot{u}_{s} + M_{tms}\ddot{u}_{tms} - \frac{\partial}{\partial x}\left(A_{tm}^{eff}G_{xy}\frac{\partial u_{tms}}{\partial x} + A_{tm}^{eff}\eta_{xy}\frac{\partial \dot{u}_{tms}}{\partial x}\right), \quad (1)$$

where $F_{hb/tms}$ is the external force (per unit length) applied by the HB of the OHC in the shear direction,[19] C_{sub}^{f} is the damping coefficient due to the viscosity of the fluid in the subtectorial space, u_s is the relative shear displacement between the TM and the reticular lamina (RL), and A_{tm}^{eff} is an effective cross-sectional area of the TM. Because of its inhomogeneity,[15] anisotropy,[14] and of its frequency dependent properties,[16,17] it is difficult to estimate the values for the effective TM shear stiffness and mass. Here we chose the TM stiffness per unit length, K_{tms}, based on the values published by Richter *et al.* for the radial TM stiffness of the gerbil (as discussed in the Appendix). Hence the TM stiffness is within one order of magnitude of the HB stiffness published in Ref. 29. For the TM shearing mass, we take into account the longitudinal variation observed in the TM cross section area in Ref. 12 and choose the value to fit the predictions of the BM response to acoustic stimulation with published experimental data. The values of the TM parameters are listed in Table I.

B. Basilar membrane mathematical model

To introduce elastic longitudinal coupling in the BM mechanics, we use an orthotropic plate model. The governing equation for the BM motion is

$$P_{bm}(x,y) = \frac{2}{b}C_{bm}\dot{u}_{bm} + M_{bm}\ddot{u}_{bm} - \frac{\partial^2}{\partial x^2}\left(D_{xx}\frac{\partial^2 u_{bm}}{\partial x^2} + D_{xy}\frac{\partial^2 u_{bm}}{\partial y^2}\right) - 2\frac{\partial^2}{\partial x\,\partial y}\left(D_{s}\frac{\partial^2 u_{bm}}{\partial x\,\partial y}\right) - \frac{\partial^2}{\partial y^2}\left(D_{yy}\frac{\partial^2 u_{bm}}{\partial y^2} + D_{xy}\frac{\partial^2 u_{bm}}{\partial x^2}\right), \quad (2)$$

where P_{bm} is the pressure applied by the fluid and the OHC on the BM,[19] C_{bm} is the BM viscous damping per unit area, M_{bm} is the mass of the BM per unit area, b is the width of the BM, and $u_{bm}(x,y)$ is the BM displacement. D_{xx}, D_{yy}, D_{xy}, and D_s are the orthotropic plate bending stiffnesses of the BM. As in Ref. 19, we assume that the BM vibrates with the mode shape $u_{bm}(x,y)=u_{bm}^{0}(x)\sin(\pi(y+\frac{b}{2})/b)$ for $-b/2\le y \le b/2$. The values of the BM parameters are listed in Table II.

C. Uncoupled structural longitudinal coupling space constants

In the macroscopic model of the cochlea, the motions of the TM and of the BM are coupled, especially in the active model. However, to evaluate the contributions of TM longi-

TABLE II. Basilar membrane properties (x is in m).

Properties	Value	Ref.
D_{xx}	0 N m in LR and TM-LC models, 6.5×10^{-11} N m in BM-LC and TMBM-LC models	24
D_{xy}	0 N m in LR and TM-LC models, 3.1×10^{-11} N m in BM-LC and TMBM-LC models	24
D_s	0 N m in LR and TM-LC models, 4.3×10^{-11} N m in BM-LC and TMBM-LC models	24
D_{yy}	$1.9\times10^{-9}\left(\frac{h_{bm}}{7\times10^{-6}}\right)$ N m	30 and 19
M_{bm}	$\rho_{bm}h_{bm}$ where $\rho_{bm}=1000$ kg/m^3	19
C_{bm}	0.85×10^{-1} N s/m^2	
b	$(80+54\times10^{-2}x)\times10^{-6}$ m	
h	$(7-2.86\times10^{-2}x)\times10^{-6}$ m	15 and 20

tudinal coupling and of BM longitudinal coupling to cochlear mechanics, we determine here the equations used to compute the space constants for the shear motion of the TM and the transverse motion of the BM uncoupled from the other structures and the fluid. Analysis of these predictions will allow us to estimate the spatial extent of a single row of OHC forcing due to longitudinal mechanical coupling in the TM and BM.

The TM is attached to the reticular lamina by three rows of HBs (each with a stiffness per unit length K_{st}) and to the spiral limbus and has a stiffness per unit length, K_{tms}. The force applied by the HB on the TM in the shear direction is given by $F_{hb/tms}=-3K_{st}u_s$. Therefore, according to Eq. (1), if the longitudinal variations in A_{tm}^{eff}, G_{xy}, and η_{xy} are neglected locally, the complex wavenumber characteristic of TM longitudinal coupling (when the TM and HB are uncoupled from the other structures), k_{tms-hb}, is given by

$$k_{tms-hb}=\left(-\frac{K_{tms}+3K_{st}-c_{tms}i\omega-M_{tms}\omega^2}{A_{tms}^{eff}(G_{xy}-i\omega\eta_{xy})}\right)^{1/2}. \quad (3)$$

If we integrate out the radial dependence of the BM displacement and neglect locally the longitudinal variations in b and of the basilar membrane properties, the BM governing equation is reduced to

$$F_{bm}(x)=C_{bm}\dot{u}_{bm}^0(x)+\frac{b}{2}M_{bm}\ddot{u}_{bm}^0(x)-\frac{b}{2}\left[D_{xx}\frac{\partial^4u_{bm}^0}{\partial x^4}-2(D_{xy}+D_s)\left(\frac{\pi}{b}\right)^2\frac{\partial^2u_{bm}^0}{\partial x^2}+D_{yy}\left(\frac{\pi}{b}\right)^4u_{bm}^0\right], \quad (4)$$

where F_{bm} is the force per unit length applied by the pressure and the OHC.[19] The complex wave number characteristic of BM longitudinal coupling (for the BM uncoupled from the organ of Corti), k_{bm}, is the solution of the following equation:

$$D_{xx}k_{bm}^4+2\left(\frac{\pi}{b}\right)^2(D_{xy}+D_s)k_{bm}^2+D_{yy}\left(\frac{\pi}{b}\right)^4-M_{bm}\omega^2-\frac{2}{b}C_{bm}i\omega=0. \quad (5)$$

The space constants characteristic of structural longitudinal coupling are then given by

$$\lambda_{tms-hb}=\frac{1}{|Imag(k_{tms-hb})|} \quad \text{and} \quad \lambda_{bm}=\frac{1}{|Imag(k_{bm})|}, \quad (6)$$

where *Imag* represents the imaginary part of the complex wave number.

D. Model activity: OHC somatic motility and HB conductance

The electrical and mechanical degrees of freedom of the model are coupled through the somatic electromotility of the OHCs and the conductance change in the HBs. HB motility is not included and the HBs are modeled as passive springs with stiffness (per unit length) K_{st} connecting the RL and the TM. Each OHC is modeled by linearized piezoelectric-like expressions relating the OHC deformation, $u_{ohc_j}^{comp}$, and fluctuating part of the transmembrane voltage, $\Delta\phi_{ohc_j}$, to the OHC force (per unit length), F_{ohc_j}, and current (per unit length), I_{ohc_j}:

$$F_{ohc_j}=K_{ohc}u_{ohc_j}^{comp}+\epsilon_3\Delta\phi_{ohc}, \quad (7)$$

$$I_{ohc_j}=\frac{\Delta\phi_{ohc}}{Z_m}-i\omega\epsilon_3u_{ohc_j}^{comp}, \quad (8)$$

where K_{ohc} is the OHC stiffness (per unit length), ϵ_3 is the electromechanical coupling coefficient (per unit length), Z_m is the basolateral impedance of the OHC, and the subscript ohc_j refers to each OHC where j corresponds to the row number, $j=1,2,3$.

The conversion of electrical power to mechanical power delivered by the OHC somatic force to the BM, $P_{ohc/bm}^{som}$, is then

$$P_{ohc/bm}^{som}=\tfrac{1}{2}\mathrm{Re}[\epsilon_3\Delta\phi_{ohc}\times v_{bm}^*], \quad (9)$$

where Re denotes the real part, v_{bm} the BM velocity, and * the complex conjugate. The value of the electromechanical coupling coefficient ϵ_3 determines the intensity of the active OHC force and we use a frequency independent value.[31]

The conductance of the HB (per unit length), G_{a_j}, is considered to change linearly with the rotation of the HB relative to the RL:

$$G_{a_j}=G_a^0+g_a\theta_{hb_j/rl_j}, \quad (10)$$

where G_a^0 is the conductance (per unit length) at the resting state, g_a represents the angular sensitivity of the mechano-electrical transducer (MET) channel (per unit length) (note that $g_a=G_a^1\times L_{hb}$, where L_{hb} is the length of the hair bundle and G_a^1 is the slope of the change in conductance with respect to the HB deflection, as defined in Ref. 19), and θ_{hb_j/rl_j} is angle of the HB relative to the RL. The data from He *et al.*[32] indicate that the transduction channel angular sensitivity is proportional to the maximum conductance and that the maximum conductance decreases from the base of the cochlea to the apex. The following spatial dependence for the MET angular sensitivity is used:

$$g_a(x)=g_a(0)e^{-\alpha x}, \quad (11)$$

where α is the spatial decay rate of the maximum conductance. $g_a(0)$ is kept as a free parameter and is considered to

TABLE III. OHC and HB properties (x is in m).

Properties	Value	Ref.
ϵ_3 for LR	$-0.616(10^{-5}+10^{-4}x)$ N/m mV	
ϵ_3 for TM-LC	$-1.04(10^{-5}+10^{-4}x)$ N/m mV	
ϵ_3 for BM-LC	$-0.784(10^{-5}+10^{-4}x)$ N/m mV	
ϵ_3 for TMBM-LC	$-1.12(10^{-5}+10^{-4}x)$ N/m mV	40
α	215 m^{-1}	32
$g_a(0)$	6.48 S/rad m for 100% activity	
K_{st}	$5.8\times10^4e^{-330x}$ N/m^2	$(K_{st})_{avg}=7.3\times10^4e^{-3.25x}$ N/m^2 in Ref. 29
L_{hb}	$(1+270x)\times10^{-6}$ m	

be independent of frequency. The nonlinearity of the conductance-deflection relationship is approximately taken into account by using larger values for $g_a(0)$ for acoustic stimulations with low sound pressure level (SPL) stimulations and a value of zero for a passive model, as discussed by Ramamoorthy *et al.*[19] The value needed to simulate (with the model, denoted as TM-LC model, that includes TM longitudinal coupling) the BM gain seen in experiment of Zheng *et al.*[33] at 10 dB SPL is denoted as 100% activity. The values of the OHC and MET channel parameters are listed in Table III.

E. Finite element solution

A full three dimensional box model solution of the cochlear would be computationally expensive. To reduce the size of the problem, modal decomposition is first used in the radial direction (y) as in Ref. 19. Three symmetric mode shapes are used for the pressure in the fluid: $\psi_n(y)=\cos(n\pi(y+w/2)/w)$ $(n=1,3,5)$ for $-w/2\leq y\leq w/2$, where w is the width of the duct. One mode shape is used for the BM transverse displacement. With this modal decomposition the three dimensional model is reduced to a series of two dimensional models which can then be post-processed to synthesize the full result. A Bubnov–Galerkin finite element method[34,35] is then used. The weak form is first derived from the strong form of the equations. In this study, a discretization of 741 nodes in the longitudinal direction (x) and 41 nodes in the z-direction was used (this was determined to be sufficiently converged for our purposes). Linear shape functions are used for the TM shear and bending displacements and for the electrical degrees of freedom. Bilinear shape functions are used for the fluid. Hermite cubic shape functions are used for the BM displacement since the governing equation [Eq. (2)] requires higher order continuity.

III. RESULTS

A. Longitudinal coupling (particularly in the TM) decreases the sharpness of the frequency response

In Fig. 2, frequency domain model predictions of the gain in BM velocity relative to the stapes' velocity at a basal location ($x=0.4$ cm) with and without longitudinal coupling in the TM and/or in the BM are compared. Four active models are used to predict the response of the BM to low level acoustic input. In the first (results shown with a thin solid line), a locally reacting model of the TM and of the BM (denoted as the LR model) is used. By this we mean that there is no longitudinal mechanical coupling in the TM or the BM (as in Ref. 19). In a model such as this, the dominant longitudinal coupling arises from two sources: fluid pressure and electrical conduction in the scalae of the cochlea. The fundamental components of the second model (denoted as the TM-LC model) are identical to the first, except that longitudinal coupling is now included in the representation of the TM according to Eq. (1) (results shown with a thick dashed line). In the third model (denoted as BM-LC model), longitudinal coupling is introduced in the BM according to Eq. (4) (results shown with a thin dashed line). In the fourth model (denoted as TMBM-LC model), longitudinal coupling is included in both the TM and the BM (results shown in thick solid line). In order to achieve the same BM gain at the characteristic frequency (CF) for the four models, the electromechanical coupling factors of the OHC [ϵ_3 in Eq. (8)] for the TM-LC, BM-LC, and TMBM-LC are, respectively, about 69%, 27%, and 79% higher than for the LR model (see Sec. IV). As shown in Fig. 2 and Table IV, the LR predicts a high mechanical quality factor ($Q_{10\ \mathrm{dB}}=15.2$). In the BM-LC model the $Q_{10\ \mathrm{dB}}$ is reduced ($Q_{10\ \mathrm{dB}}=9.2$). The TM-LC model predicts that the $Q_{10\ \mathrm{dB}}$ ($Q_{10\ \mathrm{dB}}=7.0$) is much lower than in the LR model and lower than in the BM-LC model. The $Q_{10\ \mathrm{dB}}$ predicted by the TMBM-LC ($Q_{10\ \mathrm{dB}}=5.7$) is

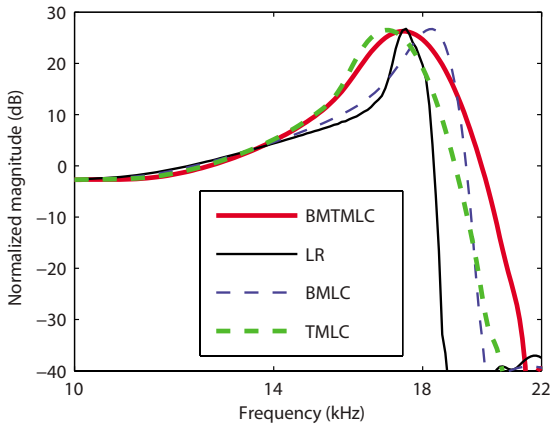


FIG. 2. (Color online) Model predictions for the BM gain in the active case (80% activity). As in Ref. 19, the gain is normalized to the maximum passive model gain. Thin solid line: Active LR model response. Thick dashed line: Active TM-LC model response. Thin dashed line: Active BM-LC model response. Thick solid line: Active TMBM-LC model response. The values for $\epsilon_3(x)$ for the different models are given in Table III. The four models are capable of predicting realistic maximum gains but at these activity levels the $Q_{10\ \mathrm{dB}}$ for the LR and BM-LC models are always much higher than that predicted by the TM-LC and TMBM-LC models.

TABLE IV. Mechanical quality factor ($Q_{10\ dB}$) and CF in the different models and in the experimental data.

Model	$Q_{10\ dB}$	CF
LR	15.2	16.7 Hz
TM-LC	7.0	16.2 kHz
BM-LC	9.2	17.5 kHz
TMBM-LC	5.7	16.6 kHz
Experimental data[a]	6.5	16.5 kHz

[a]Reference 33.

slightly lower than that by the TM-LC model. Longitudinal coupling has a small effect on the CF for a given location. Longitudinal coupling in the TM slightly reduces CF (by about 4%) while longitudinal coupling in the BM slightly increases the CF (by about 5%), as shown in Fig. 2 and in Table IV. For reference, the magnitudes of the BM gain predicted by the four models with no activity $[g_a(0)=0]$ are shown in Fig. 3. The passive response predictions are almost indistinguishable. The passive LR model is slightly more sensitive than the other models (by less than 2 dB).

The prediction for the phase of the BM relative to the stapes in the active models is shown as a function of frequency in Fig. 4. In the LR model (shown with a thin solid line) the phase accumulation at high frequency is about 8 cycles while it is only about 5–6 cycles in the other three models. The absolute value of the slope of the phase in the LR model and of the BM-LC model (shown with a thin dashed line) is higher than in the TM-LC (shown with a thick dashed line) and TMBM-LC (shown with a thick solid line) models. Hence models with longitudinal coupling in the TM have a different behavior than those with no longitudinal coupling or coupling in the BM only.

The predictions of the space constants for BM longitudinal coupling and TM longitudinal coupling [given by Eqs. (3), (5), and (6)] as a function of the frequency (normalized to CF) are shown in Fig. 5. These equations represent the effect of the two structures in isolation from the fluid and

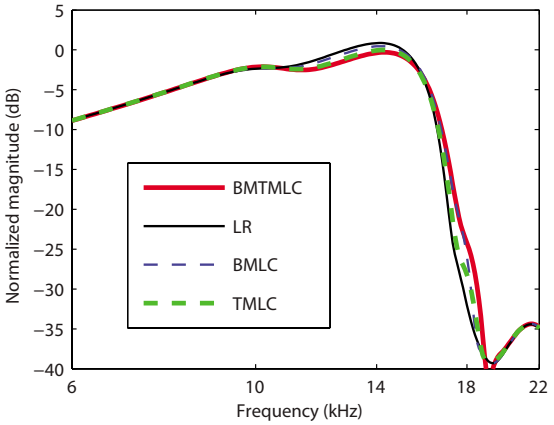


FIG. 3. (Color online) Model predictions for the BM gain in the passive case (0% activity). As in Ref. 19, the gain is normalized to the maximum passive gain. Thin solid line: Passive LR model response. Thick dashed line: Passive TM-LC model response. Thin dashed line: Passive BM-LC model response. Thick solid line: Passive TMBM-LC model response. The effect of longitudinal coupling (in the BM or in the TM) is not significant on the passive BM frequency response.

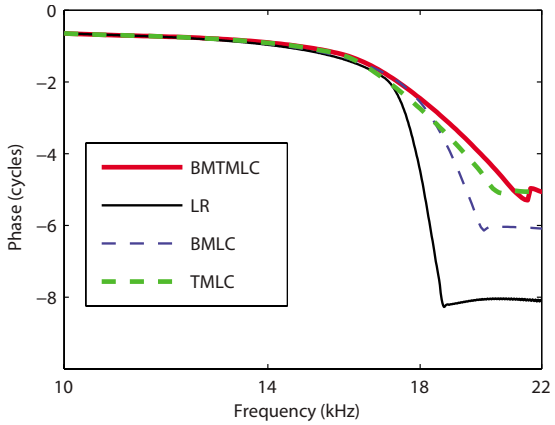


FIG. 4. (Color online) Model predictions for the BM phase relative to the stapes in the active case (80% activity). Thin solid line: Active LR model response. Thick dashed line: Active TM-LC model response. Thin dashed line: Active BM-LC model response. Thick solid line: Active TMBM-LC model response.

other structures. Even though the two constants are similar at low frequencies, the space constant characteristic of TM longitudinal coupling is more than three times higher close to CF. The TM resonance frequency is close to the CF, whereas the BM resonance frequency is much higher. Hence TM longitudinal coupling couples a higher number of OHC. Longitudinal coupling with a space constant λ can be considered significant over a distance of about 5λ.[36] Hence, at CF the TM longitudinal coupling can couple about 30 rows of OHC and the BM longitudinal coupling can only couple about 10 rows of OHCs. Therefore, as seen in Fig. 2, the presence of TM longitudinal coupling has a more significant impact on the BM response than the presence of BM longitudinal coupling.

Our results (Figs. 2, 4, and 5) show that the dominant source of structural longitudinal coupling in the cochlea is the TM viscoelasticity. Our goal is to develop a mathematical

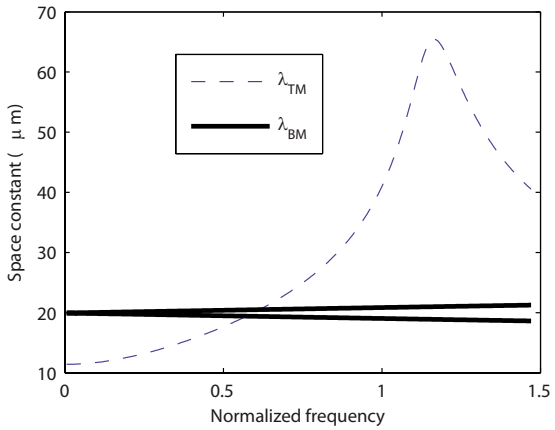


FIG. 5. (Color online) Space constants characteristic of BM longitudinal coupling (λ_{BM}) and TM longitudinal coupling (λ_{TM}) as a function of the frequency (normalized to the CF). For BM longitudinal coupling there are two space constants because Eq. (5) is quadratic in k_{bm} (Ref. 2). At low frequencies the space constants for the BM and the TM are similar (about 20 μm for the BM and 12 μm for the TM). But close to CF, the space constants characteristic of TM longitudinal coupling are more than three times as high as the space constant characteristic of BM longitudinal coupling.

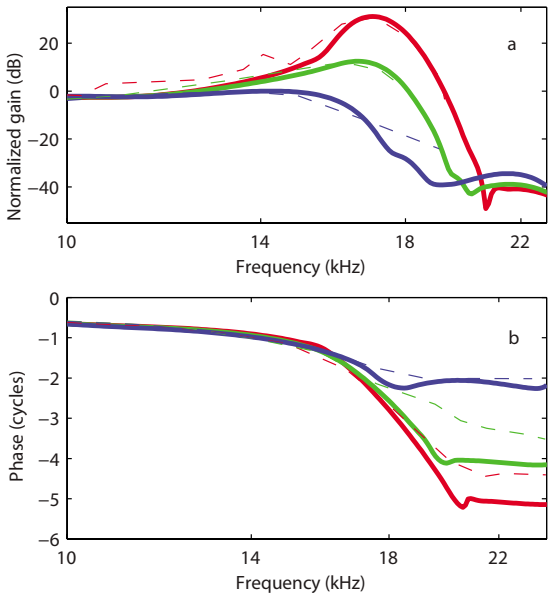


FIG. 6. (Color online) Comparison of the TM-LC model response (G_0 =7 kPa, η=0.05 Pa s) with experimental data from de Boer and Nuttall (Ref. 37). Solid lines: Model BM gain for 86% activity, 56% activity, and 0% activity. Dashed lines: Guinea pig data at 20, 80, and 100 dB. (a) Normalized (to the maximum passive response) BM magnitude in dB. (b) BM phase relative to the stapes in cycles.

model of the cochlea that is as simple as possible but can accurately predict the measurements of the BM response. Hence in the following results only the TM-LC model is used.

In Figs. 6(a) and 7, predictions of the BM gain by the TM-LC model are compared to experimental data for guinea pigs from de Boer and Nuttall[37] and Zheng *et al.*[33] Even though the protocols for these two experiments were different (in Ref. 37 bands of flat-spectrum pseudo-random noise stimulations were used while Zheng *et al.*[33] used pure tone acoustic signals), the same model is able to replicate important characteristics for both experiments. The model is a linear model and the variation from the fully active model to the passive model is achieved by decreasing the MET sensitivity $g_a(0)$ (as discussed in Ref. 19 and in Sec. II). The frequency responses of the basilar membrane track with the experimental data from de Boer and Nuttall at different SPLs [Fig. 6(a)] when the gain of the MET channels is reduced. The TM-LC model predicts around 35 dB gain for low level acoustic stimulation. As seen in experimental results, the shift in the peak frequency between the fully active and passive cases is about half an octave. The BM gain curve becomes sharper as the activity is raised. The mechanical quality factor predicted by the TM-LC model is more consistent with the experimental measurements than the results from the locally reacting model (see Fig. 2). The $Q_{10\ \mathrm{dB}}$ values predicted by the TM-LC model at the highest activity match approximately the experimental values for low SPL stimulation. For instance, $Q_{10\ \mathrm{dB}}$ is 7.7 in the fully active TM-LC model compared to 6.5 in Zheng *et al.*[33] experiment (compare the heavy solid and dashed curves in Fig. 7). The TM-LC model prediction for the BM phase follows closely the data from de Boer and Nuttall [see Fig. 6(b)]. Note that the phase data are not available from Zheng *et al.*[33] The phase accumulation at high frequency is about 5.0 cycles in the active TM-LC model compared to 4.5 cycles in the 20 dB SPL experiment. The phase accumulation at CF is about 2 cycles in both the model and the experiment. The model predictions for the phase slope at CF are slightly higher than in the experimental data.

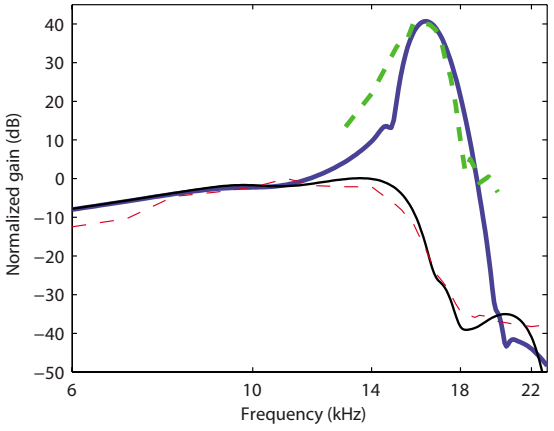


FIG. 7. (Color online) Comparison of the TM-LC model response for the gain (G_0=7 kPa, η=0.07 Pa s) with experimental data from Zheng *et al.* (Ref. 33). Solid lines: Model BM responses for 100% activity (thick line) and 0% activity (thin line). Dashed lines: Experimental data at 10 dB (thick line) and 100 dB SPL (thin line).

By taking the inverse Fourier transform of the frequency response, an impulse response can be derived from the experiments and simulations. Results from the TM-LC theory are compared to measurements from de Boer and Nuttall[37] in Fig. 8. The oscillations of the response continue up to about 1 ms in experimental data [Fig. 8(b)] and 1.4 ms [Fig. 8(a)] for the model simulations in the active case. This is a considerable improvement over the TM-LR theory, which incorrectly predicted a much longer impulse response in the active case, with oscillations that continue up to 3 ms.[19] In the passive case the oscillations are about 0.4–0.5 ms both in the experimental results and the model simulations. Note that for the passive case, the model is relatively insensitive to TM longitudinal coupling as both models match the experimental results quite well for the impulse response. The model also predicts the same zero crossings of the passive and active responses for the first few cycles, consistent with observations from de Boer and Nuttall.[37]

The simulations of the gain at different longitudinal locations of the cochlea follow the expected trend, as shown in Fig. 9. The peak of the BM gain curves shifts to lower frequency as the location approaches the apex, accompanied by a lower gain and $Q_{10\ \mathrm{dB}}$ (see also Ramamoorthy *et al.*[19]).

B. Parameter sensitivity

TM longitudinal coupling is characterized by the shear modulus (G_{xy}) and shear viscosity (η_{xy}) coefficients. Ghaffari *et al.*[20] determined the longitudinal shear modulus (G_{xy}) and shear viscosity (η_{xy}) using a mathematical model of the TM similar to what we propose and their measurements of the shear traveling wave in a mouse TM at acoustic frequencies. The results show that the shear modulus is higher at the

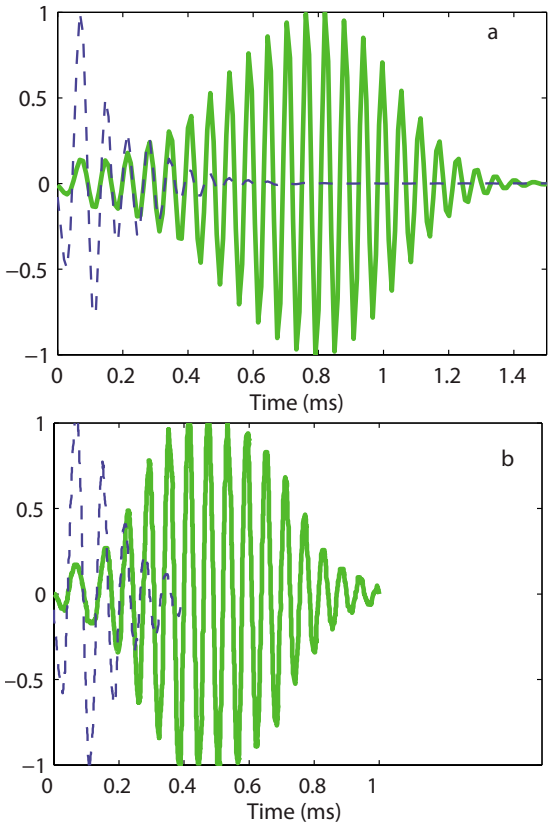


FIG. 8. (Color online) Normalized theoretical and experimental impulse response functions. (a) TM-LC model impulse response. Dashed line: Passive response. Solid line: Active response (86% activity). (b) Normalized experimental BM impulse responses from de Boer and Nuttall (Ref. 37) at 10 dB SPL (solid line) and 100 dB SPL (dashed line).

base than at the apex. Using the TM material property measurements, Gavara and Chadwick[14] estimated the shear wave velocity at a basal and more apical location in the guinea pig. Based on their value for the shear wave velocity, the shear modulus ($G_{xy}=v_s^2\rho$) is 0.67 kPa for a basal location and 0.44 kPa at a more apical location. We use for the shear modulus

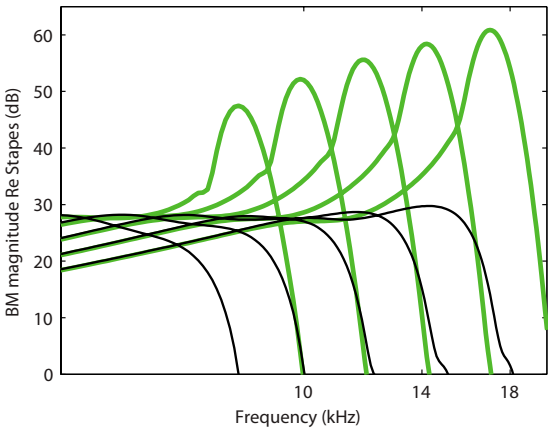


FIG. 9. (Color online) Prediction of the BM response relative to the stapes as a function of frequency in the TM-LC model at different longitudinal locations (x=0.4, 0.5, 0.6, 0.7, and 0.8 cm). The active (81% activity) model responses are shown with a thick solid line. The passive model responses are shown with a thin solid line. As the location approaches the apex, the peak of the BM gain curve shifts to a lower frequency, the magnitude of the BM gain is lower, and the tuning of the response is less sharp.

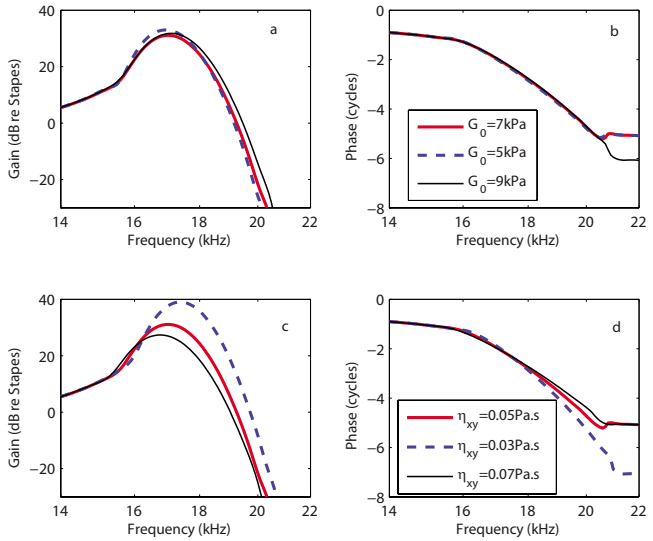


FIG. 10. (Color online) Model sensitivity to TM parameter variations. [(a) and (b)] Effect of changing the shear modulus G on the amplitude (a) and phase (b) of the BM gain relative to the stapes [legend in (b)]. [(c) and (d)] Effect of changing the shear viscosity η on the amplitude (c) and phase (d) of the BM gain relative to the stapes [legend in (d)]. For all simulations the activity is kept constant (86% activity).

the same spatial variation as for the radial stiffness of the TM [$G_{xy}(x)=G_0\exp(\alpha_{tm}x)$]. The values of the shear modulus are of same order of magnitude as the values in Ref. 14. In Ref. 20, the phase angle of the complex shear modulus at the CF of the basal location (about 80 kHz) is approximately 65°. For the guinea pig, at x=0.4 (CF=17 kHz), and if the phase angle at 17 kHz of the shear modulus is the same as in the mouse at 80 kHz, the shear viscosity should be about 0.04 Pa s.

In the theoretical results shown thus far, the shear modulus at the base is G_0=7 kPa and a constant value of 0.05 Pa s is used for the shear viscosity. Varying these parameters affects the predictions of the BM magnitude in response to acoustic stimulations only for frequencies near the CF as the results in Fig. 10 show. Increasing the shear modulus G reduces the gain at the peak location (a few dB) and makes the response less sharp, as seen in Fig. 10(a). Increasing the shear modulus is akin to increasing the longitudinal coupling of the TM; hence this result is consistent with differences seen in the TM-LC and LR models (see Fig. 2). Increasing the shear viscosity η reduces the gain, $Q_{10\ \mathrm{dB}}$, of the BM response and the phase accumulation at frequencies higher than the CF, as shown in Figs. 10(c) and 10(d). The TM shear viscosity plays an important role in modifying the slope of the phase at the CF; increasing the shear viscosity reduces the phase slope. For this range of parameters, the overall qualitative nature of the response predictions is not altered even though some of the quantitative details are affected.

IV. DISCUSSION

A. TM longitudinal coupling is necessary to predict a BM gain curve and impulse response consistent with experimental data

As shown in Figs. 2, 4, 6, and 7, only the TM-LC and the TMBM-LC models reproduce the $Q_{10\ \mathrm{dB}}$ seen in experi-

mental data for a basal location. In the model, the shear resonance of the TM (i.e., the resonance of the TM mass attached to the spiral limbus and to the HBs) occurs at a frequency close to the CF. Hence one can loosely think of the radial resonance of the TM setting the CF for a given location (in the four models). Indeed changing the TM mass directly impacts the CF, but has little effect on the passive response.[19] Somatic motility driven by the MET current is the only electromotile force in the model. For all four models it is possible to choose parameters to replicate the level of gain seen in the experimental results for low level sound [see Ramamoorthy *et al.*[19] and Fig. 6(a)]. But, as seen in Fig. 2, only the TM-LC and BMTM-LC models replicate the $Q_{10\ \mathrm{dB}}$ as well as the duration of the impulse response of the experimental data. Figure 8 shows that the duration of the impulse for the model and experiments is roughly the same. However, there are differences as the predicted maximum amplitude is reached at 0.8 ms while experimentally the peak occurs at 0.5 ms. This may be due to the slight differences in the signal processing, but more likely it is due to some deficiencies in the model (notably the nonlinearity), as minor differences in phase-frequency relations between theory and experiments are also seen. The role of TM longitudinal viscoelastic coupling is partially corroborated by the experimental results of Russell *et al.*[11] where they used a β-tectorin knockout mouse which possesses reduced longitudinal coupling (as measured by Ghaffari[10]) and showed sharper tuning, as we have predicted. Our model shows that the TM properties do not have a significant influence on the shape of the passive frequency response; such a finding is consistent with measurements of the BM response in mutant mice with a detached TM.[38] The BM-LC model predicts a $Q_{10\ \mathrm{dB}}$ that is lower than in the LR model but still much higher than in the experimental data. Since the BM resonance is at a much higher frequency than CF, the effect of longitudinal coupling in the BM equation is not as dramatic on the tuning of the BM response as the effect of longitudinal coupling in the TM. Because of the mechanical connection to the HB and the fact that our model predicts that the shear resonance of the TM corresponds to the CF at the base of the cochlea, the TM appears to be the most important structure determining the broadness of the active frequency response (at a basal location).

B. Longitudinal coupling stabilizes the linearized cochlear model

Introducing longitudinal coupling, particularly in the TM, stabilizes the linear model. For the parameters chosen in this paper, the linearized LR model is at the limit of stability for a 25 dB BM gain, whereas the linearized TM-LC model is well under the stability limit for the BM gains seen experimentally (e.g., 35 dB). The stability limit is found by increasing the MET sensitivity until the impulse response is no longer finite and bounded (i.e., the system is unstable). Numerical experiments (results not shown) indicate that this change in the stability of the system is only due to the addition of longitudinal coupling and not to the modifications of the OHC parameters between the two models. A key finding of this work is that the cochlea is stabilized, in part, by the longitudinal coupling of the TM and that this stabilization allows for the cochlea to achieve higher stable levels of gain than would be possible without the longitudinal coupling. Since we predict that the system is not near a stability limit, this suggests that spontaneous otoacoustic emissions cannot be produced by a local spontaneous oscillation but may only be explained by a global phenomenon such as the coherent wave reflection theory.[39]

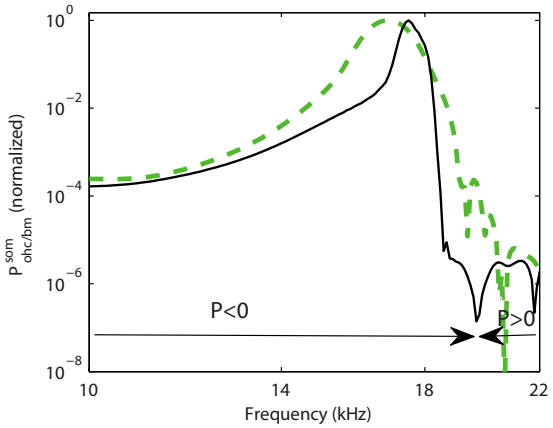


FIG. 11. (Color online) Absolute value of the power of the OHC active (somatic) force on the BM ($P^{som}_{ohc/BM}$) normalized to the value at the CF. Thick dashed line: $P^{som}_{ohc/BM}$ in the TM-LC model. Thin solid line: $P^{som}_{ohc/BM}$ in the LR model. For both models the parameters are the same as in Fig. 2. The sign of the power (P in the figure) is indicated in the figure to show the frequency region where the OHC somatic forces add power to the BM ($P^{som}_{ohc/BM}<0$) and the frequency region where it removes power from the BM ($P^{som}_{ohc/BM}>0$). For both models the OHC somatic force adds power to the BM for frequencies less than the CF and up to about 20 kHz.

When we implement our TM-LC model with a significantly reduced shear modulus and shear viscosity while keeping other parameters constant, the BM gain is higher and the cochlear model is unstable at high activity. If such an instability were present when TM longitudinal coupling is reduced, then we should expect an increase in the gain and the presence of broad band otoacoustic emissions. However, in measurements by Russell *et al.*,[11] the mutant mouse with reduced TM longitudinal coupling (as shown by Ghaffari[10]) has a reduced BM sensitivity compared to the wild-type mouse, and broad band otoacoustic emissions have not been reported. This apparent contradiction between model predictions and experimental results indicates that the mutant mouse cochlea might develop with a reduced MET sensitivity and/or OHC somatic force compared to a wild-type mouse. In order to predict the same gain at the CF, a lower value of the electromechanical coupling coefficient of the OHC [ϵ_3 in Eq. (8)] is used in the LR model than in the TM-LC model. Note that both values of ϵ_3 are realistic, within 50% of the experimental estimate from Iwasa and Adachi.[40]

C. Somatic motility of the OHC can operate at high frequencies and deliver power to the BM

Further processing of the results for the LR and TM-LC models to determine the power delivered to the BM by the OHC (see Sec. II) is shown in Fig. 11 (results shown with thin line and thick line, respectively) for a location near the

17 kHz best place. For both models, the OHCs are predicted to convert electrical to mechanical power and deliver power to the BM for frequencies lower than the CF. In the LR model, the OHCs deliver significant mechanical power to the BM in a narrower range of frequencies near the CF than in the TM-LC model (as normalized by the maximum mechanical power delivered to the system by the OHC). We attribute this to the coupling of multiple OHCs by the structural longitudinal coupling. The TM-LC and LR models predict that somatic motility of the OHC is able to deliver significant power to the BM even in the face of the RC-filtering of a 280 Hz corner frequency of the OHC at this location. The combination of a large electromechanical coupling coefficient, ϵ_3, and HB transduction current (which arises from the shear resonance of the TM at the CF) overcomes the filtering. At 86% level of activity (see Fig. 6), the prediction of the transduction current sensitivity is 2.7 nA per nm of BM deflection. The magnitude of the HB current is 1.35 nA for a 0.5 nm BM displacement (a displacement which corresponds approximately to a 20 dB SPL acoustic stimulation[41,42]). These values are lower than the maximal experimental estimates[32] (see also Ramamoorthy *et al.*[19] for further discussion).

V. Conclusions

This model demonstrates that the TM has a significant influence on cochlear tuning as the tuning of the BM response arises from an electromechanical resonance mode of the OoC controlled mostly by the TM shear mode properties. BM longitudinal coupling has a more limited impact on the BM response. Rather than intrinsically tuned electromotility or mechanotransduction (as suggested by Müller and Gillespie[43]), our results show that the sharp BM frequency response is due to the dynamics of the combined electromechanical system (which does not require the electromotility itself to be tuned). The TM is a crucial structure for active cochlear mechanics because of its connectivity to the HBs. The TM motion relative to the reticular lamina deflects the HBs and opens the transduction channels, which increases the transmembrane potential. This drives OHC somatic motility by converting stored electrical energy from the endocochlear potential to mechanical power delivered to the BM. Hence the presence of the TM and of OHC somatic motility is essential to assure a high BM sensitivity. Our model predictions for the BM frequency response and impulse response show that TM longitudinal coupling is critical for a well-functioning cochlea. The human ear needs to have both a high frequency selectivity and a short impulse response as well as a high sensitivity. This tradeoff between frequency discrimination and transient capture is controlled mostly by the longitudinal viscoelastic properties of the TM. The same longitudinal coupling is responsible for stabilizing the highly sensitive cochlea.

ACKNOWLEDGMENTS

The authors would like to thank Y. Li for her helpful discussions about the model and parameter choices. This research was supported by National Institutes of Health Grant No. NIH-NIDCD R01-04084.

APPENDIX: ESTIMATION OF THE TM STIFFNESS PER UNIT LENGTH, K_{tms}

Richter *et al.*[12] measured the TM radial stiffness in the gerbil hemicochlea using a piezoelectric probe of diameter $d=25$ μm. For static measurements, the governing equation [Eq. (1)] for the TM radial displacement of the TM is given by (neglecting the variations in A_{tm}^{eff} and G_{xy} as a function of x)

$$f_{ext}(x) = K_{tms}u_{tms} - A_{tm}^{eff}G_{xy}\frac{\partial^2 u_{tms}}{\partial x^2}, \tag{A1}$$

where f_{ext} is the force per unit length applied by the probe. The probe deforms the TM with the following mode shape:

$$u_{tms}(x) = U \quad \text{if } |x-x_0| < \frac{d}{2},$$

$$u_{tms}(x) = Ue^{-|x-x_0|/\lambda} \quad \text{if } |x-x_0| > \frac{d}{2}, \tag{A2}$$

where x_0 is the center of application of the probe force, U is the displacement of the TM at the probe tip, and $\lambda_{tm} = \sqrt{A_{tm}^{eff}G_{xy}/K_{tms}}$. Therefore the stiffness measured by the probe is

$$K = \frac{\int_{(x_0-d/2)^-}^{(x_0+d/2)^+} f_{ext}(x)dx}{U} = K_{tms}d\left(1+2\frac{\lambda_{tm}}{d}\right), \tag{A3}$$

with the parameters used in our simulations, at $x=0.4$ cm (which corresponds to the 17 kHz BP in the guinea pig), $K_{tms}=4.2$ kPa, $G_{xy}=2.1$ kPa, $A_{tm}^{eff}=3120$ μm^2 so that $\lambda_{tm}=40$ μm, and $K=4.2K_{tms}d=0.44$ N/m. In the gerbil cochlea, at the 17 kHz BP, Richter *et al.* measured $K=0.255$ N/m (when the TM is detached from the stereocilia).

[1] P. Dallos, "Overview: Cochlear neurobiology," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York, 1996), pp. 1–43.
[2] D. Strelioff, "Computer-simulation of generation and distribution of cochlear potentials," J. Acoust. Soc. Am. **54**, 620–629 (1973).
[3] W. E. Brownell, C. R. Bader, D. Bertrand, and Y. De Ribaupierre, "Evoked mechanical responses of isolated cochlear hair cells," Science **227**, 194–196 (1985).
[4] J. J. Zwislocki, "Theory of cochlear mechanics," Hear. Res. **2**, 171–182 (1980).
[5] J. J. Zwislocki, "Analysis of cochlear mechanics," Hear. Res. **22**, 155–169 (1986).
[6] J. B. Allen, "Cochlear micromechanics—A physical model of transduction," J. Acoust. Soc. Am. **68**, 1660–1670 (1980).
[7] A. W. Gummer, W. Hemmert, and H. P. Zenner, "Resonant tectorial membrane motion in the inner ear: Its crucial role in frequency tuning," Proc. Natl. Acad. Sci. U.S.A. **93**, 8727–8732 (1996).
[8] F. Mammano and R. Nobili, "Biophysics of the cochlea: Linear approximation," J. Acoust. Soc. Am. **93**, 3320–3332 (1993).
[9] R. S. Chadwick, E. K. Dimitriadis, and K. H. Iwasa, "Active control of waves in a cochlear model with subpartitions," Proc. Natl. Acad. Sci. U.S.A. **93**, 2564–2569 (1996).
[10] R. Ghaffari, "The functional role of the tectorial membrane in the cochlear mechanics," Ph.D. thesis, MIT, Cambridge, MA, 2008.
[11] I. J. Russell, P. K. Legan, V. A. Lukashkina, A. N. Lukashkin, R. J. Goodyear, and G. P. Richardson, "Sharpened cochlear tuning in a mouse with a genetically modified tectorial membrane," Nat. Neurosci. **10**, 215–223 (2007).

[12]C. P. Richter, G. Emadi, G. Getnick, A. Quesnel, and P. Dallos, "Tectorial membrane stiffness gradients," Biophys. J. **93**, 2265–2276 (2007).
[13]J. J. Zwislocki and L. K. Cefaratti, "Tectorial membrane II: Stiffness measurement in vivo," Hear. Res. **42**, 211–227 (1989).
[14]N. Gavara and R. S. Chadwick, "Collagen-based mechanical anisotropy of the tectorial membrane: Implications for inter-row coupling of outer hair cell bundles," PLoS ONE **4**, e4877 (2009).
[15]B. Shoelson, E. K. Dimitriadis, H. Cai, B. Kachar, and R. S. Chadwick, "Evidence and implications of inhomogeneity in tectorial membrane elasticity," Biophys. J. **87**, 2768–2777 (2004).
[16]C. C. Abnet and D. M. Freeman, "Deformations of the isolated mouse tectorial membrane produced by oscillatory forces," Hear. Res. **144**, 29–46 (2000).
[17]J. W. W. Gu, W. Hemmert, D. M. Freeman, and A. J. Aranyosi, "Frequency-dependent shear impedance of the tectorial membrane," Biophys. J. **95**, 2529–2538 (2008).
[18]R. Gueta, D. Barlam, R. Z. Shneck, and I. Rousso, "Measurement of the mechanical properties of isolated tectorial membrane using atomic force microscopy," Proc. Natl. Acad. Sci. U.S.A. **103**, 14790–14795 (2006).
[19]S. Ramamoorthy, N. V. Deo, and K. Grosh, "A mechano-electro-acoustical model for the cochlea: Response to acoustic stimuli," J. Acoust. Soc. Am. **121**, 2758–2773 (2007).
[20]R. Ghaffari, A. J. Aranyosi, and D. M. Freeman, "Longitudinally propagating traveling waves of the mammalian tectorial membrane," Proc. Natl. Acad. Sci. U.S.A. **104**, 16510–16515 (2007).
[21]G. V. Bekesy, *Experiments in Hearing* (McGraw-Hill, New York, 1960).
[22]L. Voldrich, "Mechanical properties of the basilar membrane," Acta Oto-Laryngol. **86**, 331–335 (1978).
[23]R. C. Naidu and D. C. Mountain, "Longitudinal coupling in the basilar membrane," J. Assoc. Res. Otolaryngol. **2**, 257–267 (2001).
[24]S. Liu and R. D. White, "Orthotropic material properties of the gerbil basilar membrane," J. Acoust. Soc. Am. **123**, 2160–2171 (2008).
[25]S. Neely and D. Kim, "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480 (1986).
[26]J. B. Allen and M. M. Sondhi, "Cochlear macromechanics: Time domain solutions," J. Acoust. Soc. Am. **66**, 123–132 (1979).
[27]C. D. Wickersberg and R. E. Geisler, "Longitudinal stiffness coupling in a 1-dimensional model of the peripheral ear," *Peripheral Auditory Mechanisms* (Springer-Verlag, Berlin, 1986), pp. 113–120.
[28]C. R. Steele and L. A. Taber, "Comparison of WKB calculations and experimental results for three-dimensional cochlear models," J. Acoust. Soc. Am. **65**, 1007–1018 (1979).
[29]D. Strelioff and A. Flock, "Stiffness of sensory-cell hair bundles in the isolated guinea-pig cochleas," Hear. Res. **15**, 19–28 (1984).
[30]A. W. Gummer, B. M. Johnstone, and N. J. Armstrong, "Direct measurement of basilar membrane stiffness in guinea pig," J. Acoust. Soc. Am. **70**, 1298–1309 (1981).
[31]G. Frank, W. Hemmert, and A. W. Gummer, "Limiting dynamics of high-frequency electromechanical transduction of outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **96**, 4420–4425 (1999).
[32]D. Z. Z. He, S. P. Jia, and P. Dallos, "Mechanoelectrical transduction of adult outer hair cells studied in a gerbil hemicochlea," Nature (London) **429**, 766–770 (2004).
[33]J. F. Zheng, N. Deo, Y. Zou, K. Grosh, and A. L. Nuttall, "Chlorpromazine alters cochlear mechanics and amplification: In vivo evidence for a role of stiffness modulation in the organ of corti," J. Neurophysiol. **97**, 994–1004 (2007).
[34]T. J. R. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis* (Dover, New York, 2000).
[35]A. A. Parthasarathi, K. Grosh, and A. L. Nuttall, "Three-dimensional numerical modeling for global cochlear dynamics," J. Acoust. Soc. Am. **107**, 474–485 (2000).
[36]R. C. Naidu and D. C. Mountain, "Basilar membrane tension calculations for the gerbil cochlea," J. Acoust. Soc. Am. **121**, 994–1002 (2007).
[37]E. de Boer and A. L. Nuttall, "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507 (2000).
[38]P. K. Legan, V. A. Lukashkina, R. J. Goodyear, M. Kossl, I. J. Russell, and G. P. Richardson, "A targeted deletion in alpha-tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback," Neuron **28**, 273–285 (2000).
[39]C. A. Shera, "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262 (2003).
[40]K. H. Iwasa and M. Adachi, "Force generation in the outer hair cell of the cochlea," Biophys. J. **73**, 546–555 (1997).
[41]N. P. Cooper, "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. (London) **509**, 277–288 (1998).
[42]A. L. Nuttall and D. Dolan, "Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556–1565 (1996).
[43]U. Müller and P. Gillespie, "Silencing the cochlear amplifier by immobilizing prestin," Neuron **58**, 299–301 (2008).

Local cochlear damage reduces local nonlinearity and decreases generator-type cochlear emissions while increasing reflector-type emissions

Wei Dong[a)]
Department of Otolaryngology, Head and Neck Surgery, Columbia University, New York, New York 10032

Elizabeth S. Olson
Department of Otolaryngology, Head and Neck Surgery, and Department of Biomedical Engineering, Columbia University, New York, New York 10032

(Received 28 September 2009; revised 18 December 2009; accepted 19 December 2009)

Distortion product otoacoustic emissions (DPOAEs) originate in cochlear nonlinearity and emerge into the ear canal as an apparent sum of emission types, one of which (generator) travels directly out and the other (reflector) travels out following linear reflection. The present study explores intracochlear sources of DPOAEs via simultaneous ear canal and intracochlear pressure measurements in gerbils. A locally damaged cochlea was produced with reduced local intracochlear nonlinearity and significant elevation of the compound action potential thresholds at frequencies represented within the damaged region. In the DPOAE the comparison of healthy to locally damaged cochleae showed the following: (1) In the broad frequency region corresponding to the locally damaged best frequency, DPOAEs evoked by wider f_2/f_1 stimuli decreased, consistent with the reduction in local nonlinearity. (2) DPOAEs evoked by narrow f_2/f_1 stimuli often had a bimodal change, decreasing in a lower frequency band and increasing in a band just adjacent and higher, and the DPOAE phase-vs-frequency slope steepened. These changes confirm the complex nature of the DPOAE. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3291682]

PACS number(s): 43.64.Kc, 43.64.Jb, 43.64.Bt [BLM] Pages: 1422–1431

I. INTRODUCTION

Otoacoustic emissions (OAEs) are one piece of evidence for the cochlea's active nonlinearity (Kemp, 1978) and have been widely used clinically for noninvasive hearing screening (Lonsbury-Martin and Martin, 1990). In a healthy ear with normal hearing thresholds, the cochlea's mechanical response is sharply tuned and compressively nonlinear (reviewed in Robles and Ruggero, 2001; Cooper, 2003). The nonlinearity is produced by outer hair cells (OHCs) (reviewed in Patuzzi, 1986). Cochlear nonlinearity generates frequency components that are not present in the stimulus. For single tones, the response includes harmonics in addition to the fundamental component (Cooper, 1998; Olson, 2004). For two tones, the response includes a family of tones at combination frequencies of the primaries, known as distortion products (DPs) and called distortion product otoacoustic emissions (DPOAEs) when measured in the ear canal. DPOAEs are known to depend on the active process in the cochlea, which is tied to the normal function of OHCs (reviewed in Dallos, 1992), even with high primary intensities (Avan *et al.*, 2003). The $2f_1-f_2$ component is the most commonly studied, is often largest, and is emphasized in this paper. When measured over a range of primary frequencies with constant primary ratio (f_2/f_1) the frequency response of a given component can be explored as a function of frequency. This "DPOAE-gram" has a complex nature, with fine structure in amplitude and a phase whose character depends on the stimulus parameters and on which DPOAE component is plotted (e.g., Mauermann *et al.*, 1999b, 1999a; Shera and Guinan, 1999; Talmadge *et al.*, 1999; Knight and Kemp, 2000). The DPOAE phase-vs-frequency can be either approximately constant or rapidly varying. These amplitude and phase characteristics can be readily explained if the DPOAE is contributed to by more than one source. Kim (1980) suggested that the DPOAE has substantial contributions from two regions, the f_2 place and the f_{dp} place. Kemp (1986) categorized these contributions as "wave-fixed" and "place-fixed" emission types, with the wave-fixed emission having a constant phase and the place-fixed emission having a rapidly varying phase. The two-source model was further developed as a two mechanism concept (Talmadge *et al.*, 1998; Shera and Guinan, 1999): The wave-fixed emission was due to nonlinear generation and the place-fixed emission to linear coherent reflection. In this conceptualization, the source of the wave-fixed or "generator" emission moved along the cochlea with approximately unchanging phase as the primary frequencies were swept at fixed ratio. The generator-type emission arose directly from the source of nonlinearity. It was expected to come from a place where both f_1 and f_2 are relatively large—which, due to the steep apical cut-off of cochlear filtering is in the region of the f_2 place. The "place-fixed" or "reflector" emission was thought to arise via linear reflection from random impedance variations, mainly within the peak of the response at the DP place (Zweig and Shera, 1995). The measured emission is the summation of generator-type and reflection-type emissions. Re-

[a)]Author to whom correspondence should be addressed. Electronic mail: wd2015@columbia.edu

 0001-4966/2010/127(3)/1422/10/$25.00

cent studies have noted a strong correlation between emission type and the primary ratio f_2/f_1. For example, human studies showed that the $2f_1-f_2$ DPOAE phase character changed systematically from rapidly varying (reflector-type emission) with primary ratio less than 1.1, to slowly varying (generator-type emission) at wider primary ratios (Knight and Kemp, 2000; Kalluri and Shera, 2001; Dhar *et al.*, 2005). Several studies have been designed to unmix the two emission types, with results that support the two-source framework. (Stover *et al.*, 1996; Talmadge *et al.*, 1999; Knight and Kemp, 2000; Kalluri and Shera, 2001; Knight and Kemp, 2001; Konrad-Martin *et al.*, 2001; Dhar *et al.*, 2005; Long *et al.*, 2008).

Numerous studies support OHC electromechanics as the fundamental source of the emissions. For example, in cochleae in which inner hair cells were selectively destroyed, the emissions were still strong (Jock *et al.*, 1996; Trautwein *et al.*, 1996). In addition, when OHCs were damaged or the endocochlear potential was reduced, emissions were reduced (Henley *et al.*, 1996; Avan *et al.*, 2003; Mills, 2003). When it comes to studies that probe the cochlear region of emission generation and thus test the emission theory, there is more uncertainty. For example, a study on the relationship between OAEs and OHC loss in damaged cochleae found that the DPOAE was insensitive to fairly extensive OHC loss (e.g., Harding *et al.*, 2002; Harding and Bohne, 2004). Other studies have suggested that the cochlear base contributes substantially to emissions, especially at moderate and high stimulus levels (Martin *et al.*, 1999; Fahey *et al.*, 2000; Martin *et al.*, 2008). The DPOAE is a simple signal emerging from a complicated system, and we still have much to learn about it.

In the present study, we developed a protocol to locally damage the cochlear partition (CP) and observed the resulting changes in DPOAEs and in the local intracochlear pressure responses measured close to the basilar membrane (BM). We made simultaneous measurements of DPOAEs and intracochlear DPs in gerbil using swept two-tone stimulation of fixed primary ratio, with the ratio either narrow (1.05) or relatively wide (1.25). The cochlear condition was evaluated by compound action potential (CAP) thresholds and single-tone intracochlear pressure responses measured at the location of damage. We found the following: (1) Local damage caused a decrease in the degree of locally measured cochlear nonlinearity upon single-tone stimulation and decrease in the intracochlear DPs upon two-tone stimulation. (2) When the primary ratio was narrow, DPOAEs had a bimodal change following damage, decreasing in a frequency band slightly lower than the damaged region and increasing within the damaged region. When the primary ratio was wider, the DPOAEs decreased when f_2 was approximately within the damaged region. (3) In the damaged cochleae the DPOAE phase tended to shift from gradually to steeply sloping in affected frequency regions when the primary ratio was narrow while it did not change substantially with wider ratio primaries. The results indicate that damage can actually lead to an increase in an emission that, based on its phase characteristics, is a reflector-type emission. In contrast, the wider ratio DPOAE, whose phase characteristics are indicative of a generator-type, was always reduced by the local damage.

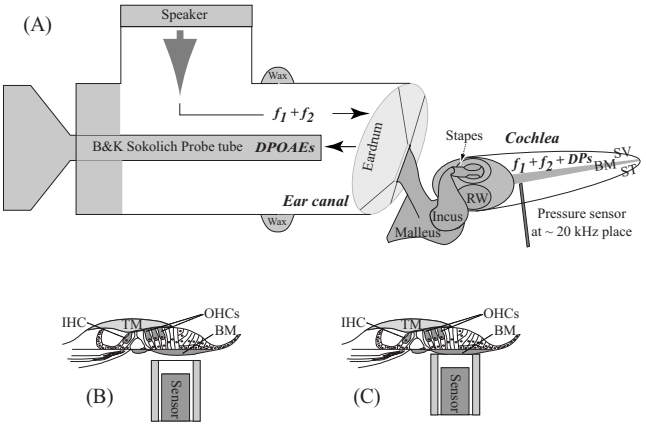


FIG. 1. Simultaneous measurement of DP and DPOAE and locally damaged cochlea. (A) DPAOE was measured in the ear canal pressure responses close to the eardrum with a sensitive microphone with two-tone stimuli to the ear canal in a closed-field sound configuration (sound system coupled to the ear canal opening with bone wax). Simultaneously, the intracochlear DP was measured in ST pressure responses in turn one with a micro-pressure-sensor. (B) Sensor positioned close to the basilar membrane. (C) Sensor indenting the cochlear partition. BM: basilar membrane; IHC: inner hair cells; OHC: outer hair cells; ST: scala tympani; SV: scala vestibuli; TM: tectorial membrane.

II. METHODS

Most of the methods have been described elsewhere (Dong and Olson, 2005b, 2006, 2008). Thus only a brief outline is given here, along with details of specific modifications.

Experiments were performed in deeply anesthetized young adult gerbils. Ketamine (40 mg/kg) was administered first to sedate the animal. Sodium pentobarbital (initial dose 60 mg/kg, supplemental 10 mg/kg) was used throughout the experiment for maintenance of anesthesia and the analgesic buprenorphine (20 mg/kg) was administered every 6 h. At the end of the experiment the animal was sacrificed with sodium pentobarbital. The care and use of the animals were approved by the Institutional Animal Care and Use Committee of Columbia University. During the experiment, animal core temperature was maintained at ~37 °C using a thermostatically controlled heating blanket. A tracheotomy was performed to maintain a patent airway. The left pinna was removed and the bulla was widely open. A small hole was hand drilled in turn one scala tympani (ST) where the best frequency (BF) is ~20 kHz to introduce the micro-pressure-sensor into the cochlea.

Acoustic intracochlear pressure and ear canal (EC) pressure responses were simultaneously recorded [Fig. 1(A)]. A sensitive Bruel & Kjaer probe tube microphone positioned within 3 mm of the tympanic membrane in a closed-field sound configuration served as the EC pressure monitor. The frequency-dependent transfer function of the probe tube microphone was accounted for when setting the sound pressure level (SPL) and analyzing the data. SPLs are reported as dB SPL (decibels relative to 20 μPa peak). Intracochlear pressure responses were measured with the micro-pressure-sensor technique (Olson, 1998). A sensor of ~125 μm outer diameter was introduced via a hand-drilled hole and positioned close to the BM at a distance of ~10 μm using a

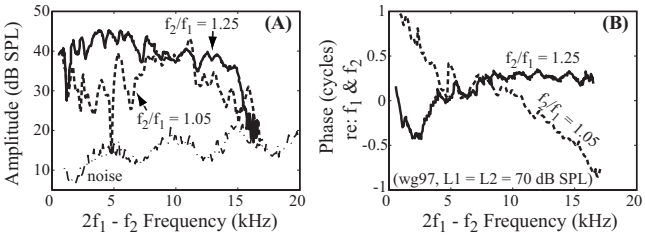


FIG. 2. Gerbil DPOAE includes multiple types. (A) Amplitude and (B) phase of DPOAE obtained from one animal (wg97, $L1=L2=70$ dB SPL, $f_2/f_1=1.05$ and 1.25, frequency step=400 Hz). Phases were normalized to ear canal f_1 and f_2 phases. Dashed and solid lines represent responses with f_2/f_1 of 1.05 and 1.25, respectively. The fine structure in the amplitude and different phase slopes indicate that the DPOAE is a combination of multiple types.

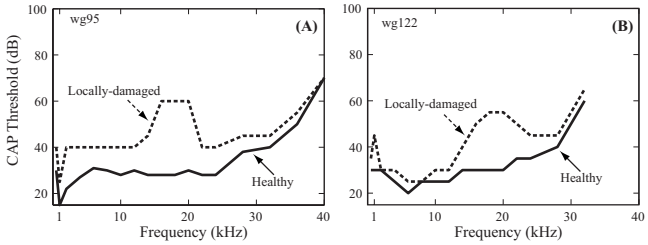


FIG. 3. CAP thresholds before and after local damage. Solid and dashed lines indicate healthy and locally damaged conditions (wg95 and wg122).

motorized micromanipulator that advanced in micrometer steps (Fig. 1). Sensor construction and calibration have been reported previously (e.g., Olson, 1998; Dong and Olson, 2006). In brief, after construction, the sensors were calibrated in water and air (air at room and body temperature) from 200 Hz to 50 kHz. The air calibration is usually similar to the water calibration. The sensitivity is typically flat with frequency to within ~2 dB through 40 kHz and often dropped a few dB between 40 and 50 kHz (Fig. 2 of Olson, 1998). Water calibration was performed before and after experiments. The sensitivity can change due to very small disturbances of the sensitive membrane; this typically produces an uncertainty of ± ~10 dB in the pressure measurements. In the present study the sensor was used to indent the cochlear partition in the turn one location of measurement [see Fig. 1(C)]. Local indentation was done in four animals. In two of these the indentation damaged the sensor's fragile membrane so that it was no longer sensitive, limiting the experiments with complete data sets pre- and post-damage to two (wg95 and wg122). A subset of results from the other two experiments is reported, and results from several other experiments are included to illustrate particular points.

The stimuli consisted of single or two-equal-intensity tones of 1–2 s duration, generated by a Tucker Davis Technology (TDT) system III and delivered via a closed system with a Radio Shack tweeter. With a fixed primary ratio paradigm, the primaries were swept in steps of ~500 Hz with f_2/f_1 fixed at either 1.05 or 1.25 (frequency spacing down to 50 Hz has been used and showed similar phase as that with step of ~500 Hz; thus there is no unwrapping phase error). The sampling frequency of the TDT system was 200 kHz. Stimulus and acquisition programs were written in MATLAB and the TDT visual design studio. Responses were later analyzed by Fourier transform with MATLAB. The phases of the DP or DPOAE are presented referenced to the EC primary phases: $\Phi_{2f1-f2\text{-re-EC}}=\Phi_{2f1-f2}-(2\Phi_{f1}-\Phi_{f2})$, where Φ_{f1} and Φ_{f2} are the phases of EC primaries.

With 1 s data acquisition time, the microphone noise level (with probe tube) was ~5–10 dB SPL up to 30 kHz and higher at higher frequencies. The noise level of the intracochlear pressure measurement is 50–60 dB SPL. System distortion has been discussed previously (Dong and Olson, 2006, 2008). Acoustic two-tone distortion was checked in a small cavity and was ~70 dB below the primary levels when both primaries were 100 dB SPL. System distortion was also checked with postmortem responses after each *in vivo* experiment.

Single- and two-tone responses were collected consecutively with the sensor close to the BM under normal condition [Fig. 1(B)] and locally damaged condition [Fig. 1(C)]. The locally damaged cochlea was produced by indenting the cochlear partition ~10 μm several times with the micropressure-sensor [Fig. 1(C))]. The physiological condition of the cochlea was monitored using CAP threshold responses (threshold criterion ~5 μV p-p) to tone pips, measured with a silver electrode on the bone surrounding the round window (RW) opening. The CAP thresholds were measured a few times during the experiments, especially before and after introducing the sensor into the ST and after damaging the CP. Damage was confirmed by an elevation in CAP thresholds. We have shown that the mere presence of the sensor close to the BM has minimal effect on CAP thresholds and DPOAEs (Dong and Olson, 2008).

III. RESULTS

A. Basic characteristics of DPOAE—Evidence for multiple types

Typical $2f_1-f_2$ DPOAE data from an intact gerbil cochlea are shown in Fig. 2. The primary levels were $L_1=L_2=70$ dB SPL, and the primary ratio (f_2/f_1) was fixed at either 1.05 (dashed line) or 1.25 (solid line). DPOAEs are robust in gerbils, and their levels could exceed 40 dB SPL at both ratios (Mills, 2002). Above 15 kHz, the DPOAEs diminished, sometimes slightly and other times more acutely; this cut-off frequency varied across animals, similar to the observed variation in CAP threshold at frequencies above 20 kHz. Deep notches in the amplitude are suggestive of cancellation, and the amplitude fine structure is one of the observations leading to a two-source emission framework. The phase was quite flat for the wider ratio DPOAE (1.25) consistent with past observations (Dong and Olson, 2008), while the narrow ratio DPOAE (1.05) had a more rapidly varying phase. These characteristics are consistent with DPOAE observations in the literature discussed in the Introduction.

B. Changes post-damage

The impact of the local damage was assessed using CAP threshold responses to tone pips (Fig. 3) and single-tone in-

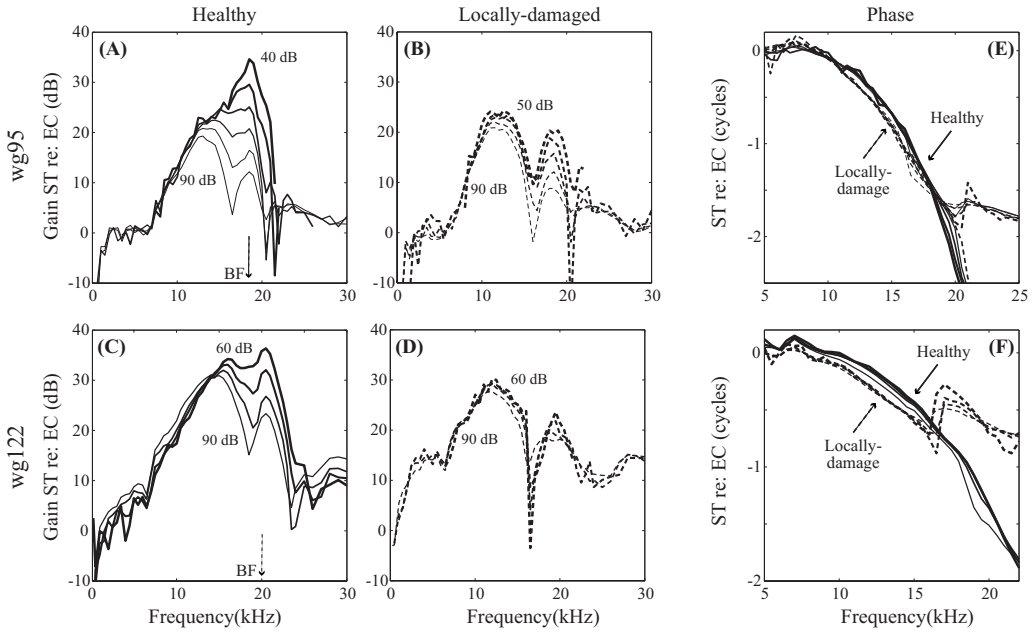


FIG. 4. Cochlear mechanics before and after local damage. Single-tone intracochlear pressure responses measured at the same location, close to the BM (~10 μm) pre- and post-damage in two animals (wg95 and wg122). [(A) and (C)] Pre-damage gain relative to stimulus pressure in the ear canal; [(B) and (D)] Post-damage gain. Stimulus SPLs are labeled. [(E) and (F)] Single-tone pressure phase referenced to EC. Solid and dashed lines represent the cochlear conditions, healthy or locally damaged, respectively.

tracochlear pressure responses measured close to the BM (Fig. 4).

1. CAP thresholds after local damage

CAP thresholds from 0.5 to 40 kHz under healthy (pre-damaged) and locally damaged (post-damage) conditions in two animals are illustrated in Fig. 3. The CAP thresholds in the healthy condition were ~30 dB SPL at frequencies below 20 kHz, and higher at higher frequencies (solid line in Fig. 3). Making the ST hole and introducing the sensor to a position close to the BM changed the CAP thresholds very little (e.g., Fig. 4 in Dong and Olson, 2008). After using the sensor to indent the cochlear partition as illustrated in Fig. 1(C), the CAP thresholds were elevated up to 20–30 dB at frequencies around 20 kHz, the BF place of the sensor, and ~10 dB or less at lower and higher frequencies (dashed line in Fig. 3). The range of substantial (>10 dB) threshold elevation is between ~12 and 24 kHz in both preparations. Thus, indenting the cochlear partition from the BM side produced a pronounced elevation in CAP threshold in the ~20 kHz BF region where the probe was positioned, leading to a locally damaged condition.

2. Single-tone responses in locally damaged cochleae

Pre- and post-damage intracochlear pressure responses to single tones can be seen in Fig. 4. The sensor was positioned close to the BM in turn one of the gerbil cochlea, where the BF is ~20 kHz. As shown in the normalized results of Figs. 4(A) and 4(C), the responses were compressively nonlinear at frequencies around the BF (19 and 20 kHz, respectively) in the healthy condition. At frequencies more than approximately a half octave (factor of 0.7) below the BF the pressure responses overlaid each other, indicating linear scaling for these frequencies. Above that frequency there is a dramatic fanning out of the normalized responses in the peak region. In the plateau region above the peak, responses typically scale linearly as in Fig. 4(A). After indenting the cochlear partition with the sensor ~10 μm, the degree of nonlinearity in single-tone responses was diminished [Figs. 4(B) and 4(D)]. On the other hand, the responses at frequencies well below and well above the BF changed little. Therefore, the local damage to the cochlear partition led to a decrease in the degree of nonlinearity at frequencies around the BF, with little change in the responses at frequencies above and below.

The phase responses in Figs. 4(E) and 4(F) (ST pressure re: EC pressure) show traveling wave delay through ~2 cycles. Post-damage, two changes occur. The high frequency phase plateau is encountered at lower frequencies, signifying that the fast, compression wave dominates the traveling wave at lower frequencies, which is to be expected since the nonlinear traveling wave response mode is diminished by the damage, but the linear fast response mode is not. (See our previous papers, Olson, 1999, 2001and Dong and Olson, 2005b for background on the multi-mode nature of intracochlear pressure.) A second change is that within the traveling wave region the phase accumulates more rapidly in the damaged preparation. This is consistent with reduced CP stiffness (Kolston, 2000). Thus the local damage has reduced the local cochlear nonlinearity, and also seems to have produced a reduction in the local stiffness of the CP.

3. Intracochlear DP and DPOAE of $2f_1 - f_2$

Simultaneous measurements of DP and DPOAE were obtained under normal and locally damaged conditions with f_2/f_1 fixed at 1.25 (Fig. 5) and 1.05 (Fig. 6). The micro-pressure-sensor was positioned ~10 μm from the BM. The

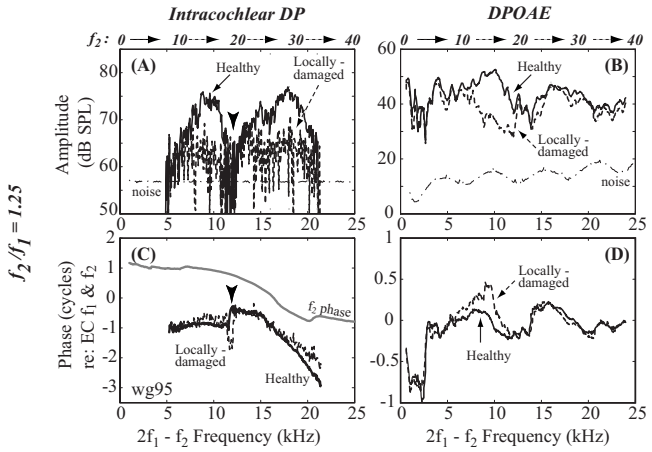


FIG. 5. $2f_1-f_2$ with $f_2/f_1=1.25$ in the cochlea (DP) and ear canal (DPOAE) pre- and post-damage. (A) DP amplitude, measured at a distance ~10 μm from the BM; (B) DPOAE amplitude; (C) DP phase; and (D) DPOAE phase. All phases are referenced to EC f_1 and f_2 phases. Solid and dashed lines represent cochlear condition, healthy or locally damaged. Gray line in panel (C) shows f_2 phase. There was no change in the low frequency region below 6 kHz (wg95, $L_1=L_2=80$ dB SPL).

response amplitudes and phases are plotted vs their own frequencies. Phases are referenced to EC f_1 and f_2 phases.

When f_2/f_1 was fixed at 1.25, DP in the healthy cochlea showed fine structure superposed on double peaks separated by a deep notch [solid line in Fig. 5(A)]. One peak reached its maximum at ~19 kHz BF and the other at ~9 kHz. The phase [solid line in Fig. 5(C)] changed character at the notch frequency [arrowhead in Fig. 5(C)]. At frequencies around the BF the DP phase was similar to that of f_2 [gray line in Fig. 5(C)], suggesting that the DP in this region was dominated by forward going distortion originating from more basal locations (Ren, 2004; Dong and Olson, 2005b; de Boer *et al.*, 2008; Dong and Olson, 2008). In the frequency region well below the BF, the DP phase departed from the f_2 phase and was more similar to the DPOAE phase [solid in Fig. 5(D)], suggesting that the DP was a combination of locally generated distortion, and apically generated distortion that was recorded on its way out of the cochlea (Dong and Olson, 2008). The f_2 frequencies are indicated above the panels, and when combined with the results in Fig. 4, help to guide expectations on whether local nonlinearity probably contributed to those DPs, or whether they were probably produced remotely. Examination of Fig. 4(A) reveals that ~12 kHz is the lowest frequency with substantial nonlinearity in single-tone responses. A 12 kHz f_2 corresponds to a 7–8 kHz $2f_1-f_2$ DP. Thus, the lower peak in Fig. 5(A), which extends from 5 to 12 kHz, is probably a combination of both apically and locally generated nonlinear distortions. The pronounced notch in the amplitude at ~12 kHz corresponded to a 180° phase shift [arrowheads in Figs. 5(A) and 5(C)], and appeared to result from cancellation of anti-phase components, probably a combination of basally generated/forward traveling, apically generated/backward traveling, and locally generated components. This notch was not seen in the DPOAE [see Figs. 5(B) and 5(D)], which reinforces that the locally measured DP has a different composition than the DPOAE. Post-damage, the amplitude of the DP decreased in both peaks. The lower frequency peak was greatly diminished above 7 kHz, but below 6 kHz the DP was little affected by the damage, consistent with the DP at these frequencies arising from more apical, undamaged regions. The phase changed relatively little [dashed lines in Fig. 5(C)] and still underwent a change in character at the notch frequency, which shifted up slightly in frequency.

In the healthy condition, the DPOAE was ~40 dB SPL up to 24 kHz with fine structure in amplitude [solid line in Fig. 5(B)] and phase varying little and nonmonotonically with frequency [solid line in Fig. 5(D)]. After the cochlea was locally damaged the DPOAE changed. In the frequency region from ~7 to 12 kHz, the DPOAE decreased, similar to the DP in that region [dashed line in Fig. 5(B)]. The small change in phase was similar in the DP and DPOAE. However, for DPOAEs with frequencies within the damaged region, the pre- and post-damage DPOAEs were similar, both in amplitude and phase. It appeared that the decrease in DP had little effect on the DPOAE in this frequency region, indicating that the DPOAE did not arise from its own best place. The relationship between local damage (as indicated in CAP changes) and DPOAE changes will be summarized in Fig. 8.

When the primary ratio was 1.05, the changes in the locally damaged condition were different (Fig. 6). The pre-damage DP amplitude had similar tuning as the single-tone response (Fig. 4), but with more fine structure [solid line in Fig. 6(A)]. At 12 kHz, there was a shallow notch in the amplitude corresponding to an inflection in the phase slope [arrowheads in Figs. 6(A) and 6(C)]. At frequencies above 12 kHz, the DP peaked at the BF (19 kHz) and its phase was similar to that of f_2 [gray line in Fig. 6(C)]. The DP tuning and its phase-vs-frequency slope at frequencies around the BF suggested that the DP was largely locally generated or basally generated/forward going. At frequencies below 12 kHz, the well-sub-BF region, the phase differed from that of the forward f_2 phase, which suggested that in this region the DP was not locally generated/forward traveling and likely originated from a more apical place (Dong and Olson, 2008). After locally damaging the cochlea, the DP in the sub-BF region decreased around 10 kHz, and the phase steepened markedly [dashed line in Fig. 6(C)]. The steepened phase

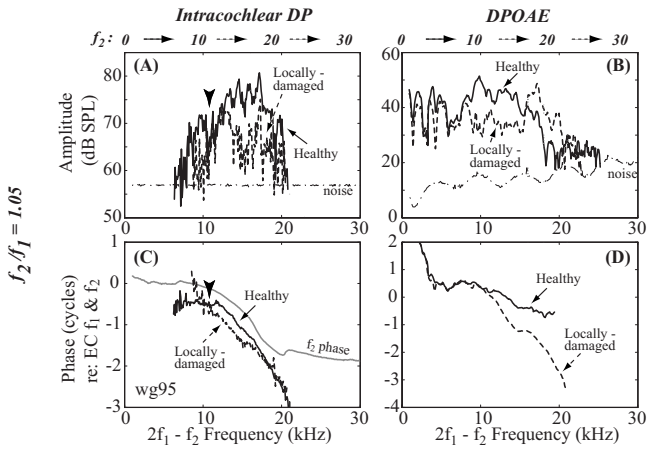


FIG. 6. $2f_1-f_2$ with $f_2/f_1=1.05$ in the cochlea (DP) and ear canal (DPOAE) pre- and post-damage. Same format as Fig. 5 (wg95, $L_1=L_2=80$ dB SPL).

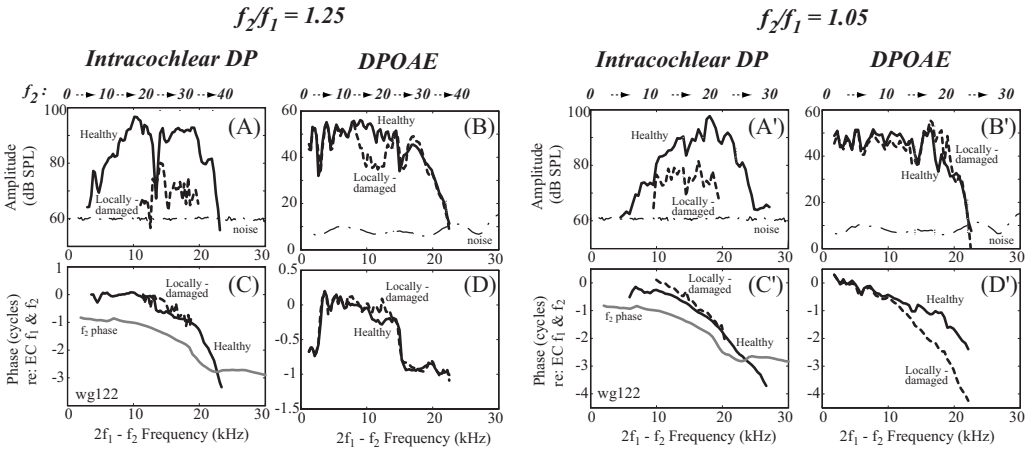


FIG. 7. $2f_1-f_2$ DP and DPOAE of animal wg122. DP was measured at ~10 μm from the BM pre- and post-damage ($L_1=L_2=80$ dB SPL, f_2 frequency step was 0.5 kHz). Same format as Fig. 5.

may indicate that the locally generated/forward going type of the DP was diminished relative to the component that was traveling out of the cochlea from more apical regions. At frequencies above 12 kHz the amplitude of the DP decreased and its phase changed slightly, but maintained its forward going character.

Pre- and post-damage DPOAEs are shown in Figs. 6(B) and 6(D). Pre-damage the DPOAE was at a level of ~40 dB SPL, with fine structure and a sharp decrease for frequencies above ~18 kHz. The phase had a fairly shallow slope, accumulating about 1 cycle over 20 kHz. There was a stair-step pattern superimposed on the shallow slope (see Talmadge *et al.*, 1999), suggesting that the emission had contributions from components with both rapidly varying and slowly varying phase-vs-frequency slopes. Following damage, at frequencies from ~9 to 17 kHz (just below the damaged region) the DPOAE amplitude decreased ~10 dB and the phase steepened substantially. At frequencies in the damaged region (~17–20 kHz) the DPOAE amplitude *increased* and the phase steepened.

The variations in DP and DPOAE after local damage were confirmed by results in a second animal (wg122) shown in Fig. 7. When the wider ratio was examined the changes are very similar to those in Fig. 5. With the narrow ratio, the DPOAE changes are more subtle, but qualitatively similar to Fig. 6.

To summarize the findings, local damage reduced cochlear nonlinearity both in single-tone responses and in the locally measured DP. With primary ratio of 1.25, the DPOAE at frequencies slightly lower than the BFs of the damaged region decreased and changed little for frequencies close to the BFs of the damaged region, with little variation in phase. With relatively narrow primary ratio (1.05), the DPOAE decreased at frequencies slightly lower than the BFs of the damaged region, and increased at frequencies close to the BFs of the damaged region. The phase changed from slowly varying to rapidly varying with frequency.

4. Correlation between DPOAE and cochlear condition

A full understanding of the changes seen benefits from a direct comparison of changes in both CAP thresholds and DPOAE amplitudes. Using the results from the two animals that were shown in Figs. 3 and 5–7, the level shifts of CAP threshold (pre-damage—post-damage) and DPOAE (post-damage—pre-damage) upon local damage are plotted together in Figs. 8(A), 8(B), 8(E), and 8(F). The DPOAE variations upon damage from two additional local damage experiments are plotted in the other four panels. The sensor was broken and CAP data were not gathered post-damage in these animals but in all cases the sensor, which was used to

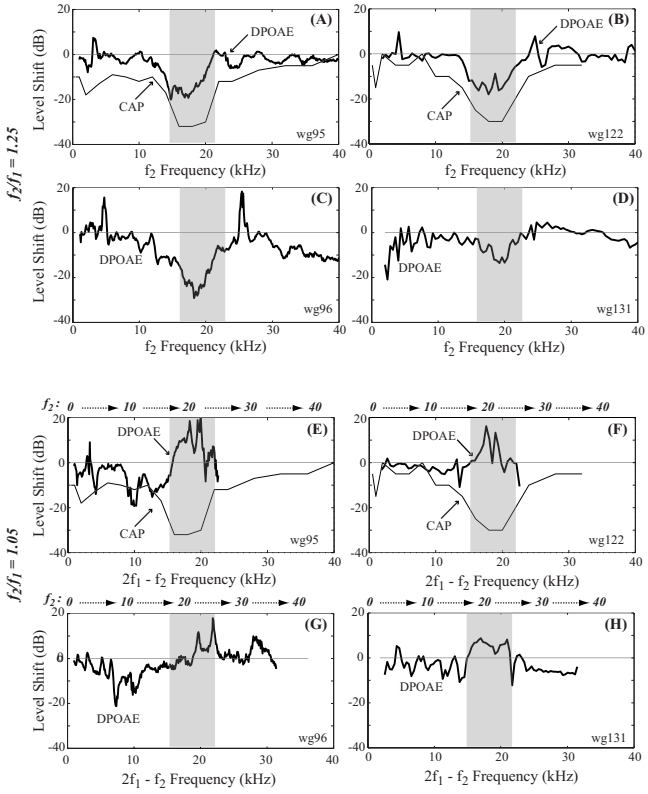


FIG. 8. CAP threshold shifts and changes in $2f_1-f_2$ DPOAE level upon local damage. Thick and thin lines show changes in DPOAE and CAP thresholds upon damage. [(A)–(D)] $f_2/f_1=1.25$, plotted vs f_2; [(E)–(H)] $f_2/f_1=1.05$ plotted vs $2f_1-f_2$ frequency. Data were from locally damaged experiments wg95, wg96, wg122, and wg131. Gray bands show the damaged region based on the sensor location (~20 kHz region) and verified with CAP changes in animals wg95 and wg122.

cause damage, was positioned at the region with ~20 kHz BF, indicated by gray bands. The DPOAE results are supportive of the results of the two main animals. The level shifts are plotted vs f_2 when the primary ratio was 1.25, and vs $2f_1-f_2$ when the primary ratio was 1.05, because these respective correlations were more instructive. (The f_2 frequency is included at the top of each panel for reference.)

When the primary ratio was 1.25 [Figs. 8(A)–8(D)], the reduction in DPOAE lined up with the damaged region (~12–24 kHz according to CAP changes) with the maximum increase in CAP thresholds. Thus, when the f_2 region was damaged, the DPOAE was reduced. With narrower primary ratio of 1.05, DPOAE *enhancement* lined up with the damaged area indicated by the CAP threshold shift [Figs. 8(E) and 8(F)]. A reduction in DPOAE level occurred at slightly lower frequencies.

IV. DISCUSSION

The present study used localized intracochlear damage, localized measurements of intracochlear responses, and emission and CAP measures to explore the two-source model and more general questions of where emissions come from in the cochlea. The present study advances these questions with a well controlled damage location, and by probing the changes in intracochlear mechanics precisely at that location, while correlating the intracochlear measurements with DPOAE and CAP.

Our research used relatively high sound pressure levels. There is strong evidence for the two-source model at low sound pressure levels, where cochlear excitation patterns are expected to be fairly well localized (Mauermann *et al.*, 1999b, 1999a; Kalluri and Shera, 2001). At moderate to high sound pressure levels the cochlear regions responsible for the emission appear to broaden due to the broadening of cochlear tuning with level (e.g., Knight and Kemp, 2001; Zhang and Mountain, 2008; Martin *et al.*, 2008). In studies in humans, at moderate to high sound levels, an emission with rapidly varying phase was observed (Martin *et al.*, 2008). Based on this phase character, the two-source model would categorize the emission as a reflector-type, and predict that it arose from reflection from around the f_{dp} place. However, the suppression behavior of the emission indicated that it arose from basal to the f_2 region. This suggests that the two-source model should be used with caution at moderate-high levels. The present study advances the understanding of DPOAE sources at relatively high primary levels, probing the limitations of the two-source model, and hopefully facilitating the use of high levels in the clinic, where high levels are often required to elicit emissions in impaired ears.

Previous studies have demonstrated the qualitative differences between emissions produced with narrow and relatively wide ratios, with narrow ratios giving rise to emissions with rapid phase variations, thought to be reflector-type emissions, and ratios above ~1.2 giving rise to emissions possessing slow phase variations, thought to be generator-type emissions (Knight and Kemp, 2000; Kalluri and Shera, 2001; Knight and Kemp, 2001; Dhar *et al.*, 2005). This dependence of emission type on ratio is expected when the phase interferences of the spatially disparate sources are taken into account (de Boer *et al.*, 2005; Shera and Guinan, 2007a, 2007b). Our own previous work also showed this dichotomy, with the wider primary ratio giving rise to a generator-type emission, and the narrow primary ratio giving rise to a mixture of emission types (Dong and Olson, 2008). The present study supports the dichotomy.

A. Relationship between DPs and DPOAEs, regions that source the DPOAEs

The relationship between the intracochlear DP at one site in the cochlea, and the DPOAE was explored in a previous publication from our laboratory (Dong and Olson, 2008). A short review follows. After generation, a DP is subject to cochlear filtering and the measured DP frequency response is determined by both the tuning of the primaries and the subsequent tuning of the DP (Dong and Olson, 2005b, 2005a). As illustrated in Fig. 5 (wider primary ratio) and Fig. 6 (narrow primary ratio), *at frequencies* approximately the BF, DP amplitude tuning and phase are similar to those of the primaries (Cooper and Rhode, 1997; Robles *et al.*, 1997; Dong and Olson, 2005b, 2005a, 2008). The DP appears to be dominated by local generation and/or forward traveling, basally generated distortion [Figs. 5(C), 6(C), 7(C), and 7(C′)]. *At frequencies well below* the BF, DP behavior is less primary-like. The DP normally shows more fine structure, similar to that in the DPOAE, and its phase departs from that of the primaries and is more like that of the DPOAE (e.g., Fig. 9 in Dong and Olson, 2008 and Fig. 8 in Rhode, 2007). Thus, at low frequencies, the DP appears to have a substantial contribution from apically generated distortion [Figs. 5(C) and 6(C)], traveling out of the cochlea. These basic observations underscore the fact that, except at primary frequencies substantially lower than the local BF, the *local* DP appears to be mainly locally generated or forward traveling and thus is not tightly linked to the DPOAE.

The results after local damage in the present paper are consistent with this picture. We have noted that at the wider primary ratio (1.25), the $2f_1-f_2$ emission appears to be dominated by the generator-type, and consider this case first. When the intracochlear probe and damage location (~20 kHz place) was at the f_2 place or somewhat basal to the f_2 place (for example, f_2 from 12 to 20 kHz in Fig. 5) both the DP and the DPOAE were reduced (Figs. 5 and 7). When f_2 was less than 12 kHz (a situation with excitation patterns that peak substantially apically of the damage location) the emission was not affected. From this result it is clear that the emission source is localized to a region that includes the f_2 place and extends somewhat basal (almost an octave). When the damaged location was somewhat apical to the f_2 place, the intracochlear DP could be much reduced with little or no change in the DPOAE (Fig. 5, and Fig. 7 unprimed panels). Thus, regions apical to the f_2 place appear to be not at all responsible for the generator-type emissions. This result makes sense, given the sharp apical drop off of the f_2 response beyond its best place. Thus, these damage studies show that substantial DPs exist within the cochlea

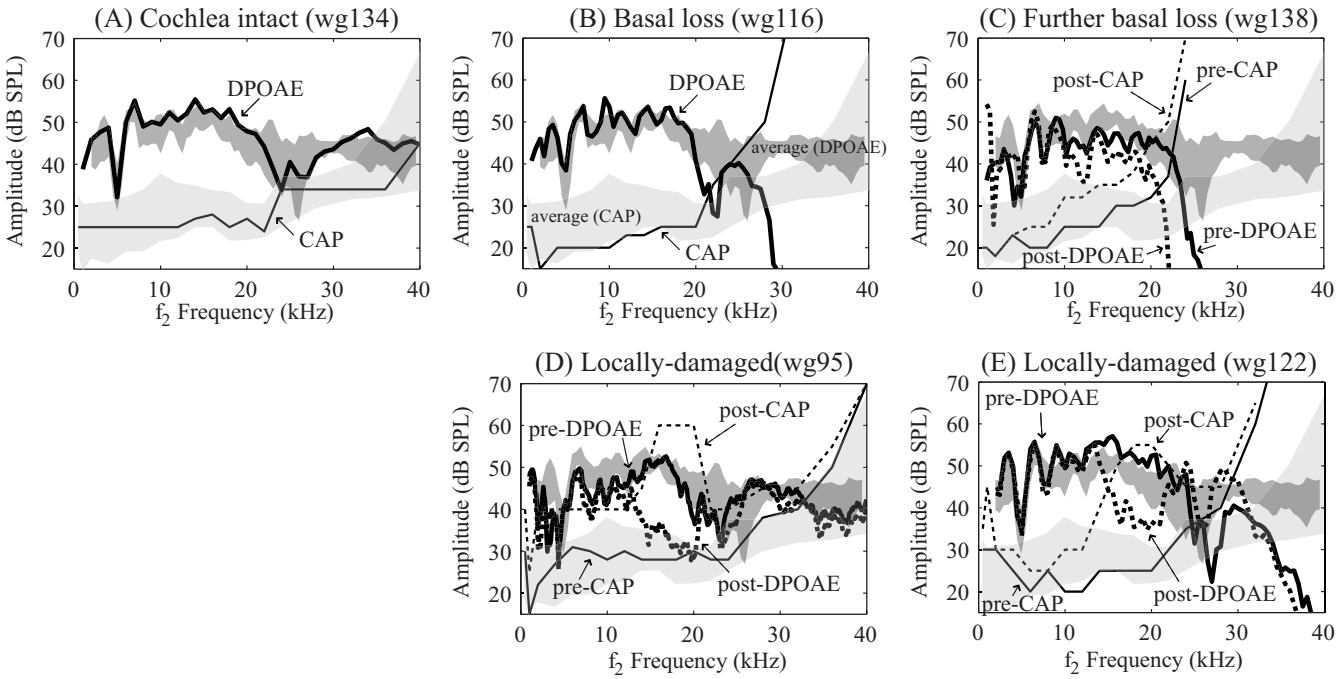


FIG. 9. Using DPOAE to gauge the cochlear condition (five animals). The darker gray bands display average DPOAE and the lighter gray bands the average CAP thresholds across eight animals with normal CAP thresholds. The width of the bands indicates the standard deviation. The lines show DPOAE and CAP results from individual animals. The thicker lines show DPOAE levels, with the solid line pre-damage and the dotted line post-damage. The thinner lines show the corresponding CAP thresholds curves, solid pre-damage, and dotted post-damage. The DPOAE shown is at frequency $2f_1-f_2$ with $L_1=L_2=80$ dB, $f_2/f_1=1.25$.

that are not detected in the emission—most certainly because they are traveling forward to be absorbed and not outward to become an emission.

With narrow primary ratio the mixture of emission types in the DPOAE is pronounced. In this case the DPOAE increased at frequencies in the damaged region, and decreased at frequencies slightly lower. Increases in DPOAEs after sound-induced cochlear lesions near the base have been reported previously (Kakigi *et al.*, 1998b, 1998a). Based on the steepened phase post-damage, the changes (Figs. 6 and 7 primed) are interpretable as a relative increase in reflector-type emission and decrease in generator-type emission. With narrow ratio primaries the physical sites of generation and reflection emissions are not well separated, and the sizes of generator- and reflector-type of emissions in the healthy cochlea appear to be similar. The damaged region appears to cause a pronounced reflection. In the case of wider primary ratio of 1.25, there is no evidence for enhanced reflection—the emission goes down, and based on its phase, remains dominantly a generator-type emission.

B. Cochlear damage changed the relative contribution from different emission types

1. Relatively wide f_2/f_1 ratio emissions can be used to gauge cochlear condition

The flat phase-vs-frequency response of the DPOAE observed when the primary ratio is fixed is the signature of a generator-type emission (Shera and Guinan, 1999). This occurs due to approximate "scaling symmetry," which refers to the way the response pattern (amplitude and phase) shifts along the cochlea but does not change shape when primary frequencies are swept at constant ratio. The generator phase is therefore constant, giving rise to a flat phase-vs-frequency response in the emission. Consistent with the findings of others, we found that the generator-type emission dominated the DPOAE at the wider primary ratio in gerbils. We also found that when the region corresponding to f_2 and basal—the putative generator region—was damaged, the DPOAE was reduced. This is illustrated in Fig. 9, which shows the covariance of the DPOAEs and CAP thresholds. The averaged DPOAE ± standard deviation from eight healthy ears is shown as the dark-gray band in each panel (stimulus level $L_1=L_2=80$ dB SPL, $f_2/f_1=1.25$). Cochlear health was determined with CAP: When the cochlear condition was good, CAP thresholds were less than 40 dB up to 35 kHz. The CAP ± standard deviation of these animals is shown in the light gray band. In Fig. 9(A) results from a healthy cochlea are shown, with CAP the thin line and DPOAE the thick line. The DPOAEs were around 40 dB and could be recorded up to 40 kHz, and were similar to the averaged DPOAEs from healthy ears. Some animals had high frequency hearing loss at the start of measurement, caused by inadvertent surgical trauma or cooling upon opening the bulla, without or prior to purposeful local damage [Figs. 9(B), 9(C), and 9(E)]. In these cases the DPOAE dropped off sharply above 25–30 kHz, and CAP thresholds increased sharply at these frequencies. Figure 9(C) documents increasing damage through two stages, indicated with solid and dotted lines (thick for DPOAE and thin for CAP). Figures 9(D) and 9(E) repeat the information in Fig. 8 showing the close correspondence between CAP threshold elevation and DPOAE decrease following local damage. These results indicate that the DPOAE with wider primary ratio can be used to gauge the cochlear

condition in the f_2 region. Looking at the DPOAE data from the panels in Fig. 9 with a "clinical" eye, it would be easy to pick out a cochlea with broad basal damage of unknown etiology [Figs. 9(B), 9(C), and 9(E)] compared to the healthy/normal DPOAE in the "healthy" ears in the dark-gray band. On the other hand, the bounded dips in DPOAE in Figs. 9(D) and 9(E) would be easy to pick out only if the pre-damage DPOAE data were also available from the same individuals. If only the post-damage data were available the dips observed would point to a region of possible cochlear damage to be followed up with further testing.

2. Changes in emission and intracochlear phase confirm changes in the dominance of the two emission types with narrow ratio

Steep phase-vs-frequency in the DPOAE at fixed f_2/f_1 is the signature of dominance by reflector-type emissions (Stover *et al.*, 1996; Shera and Guinan, 1999). In the two-source theory, this emission type arises due to linear coherent reflection from random perturbations along the cochlear partition (Zweig and Shera, 1995). The largest reflection is expected to be from the region where the traveling wave peaks, where the phase changes rapidly with frequency. On average, the phase-vs-frequency slope of the emission is predicted to be slightly less than twice the slope of the forward wave, due to the round trip travel in and out of the cochlea (Shera *et al.*, 2008). With narrow primary ratios, the DPOAE phase at low SPL normally varies rapidly with frequency in humans (Knight and Kemp, 2000, 2001) and other mammals (e.g., Withnell *et al.*, 2003) [and even non-mammals, but this is not the focus of the present paper (e.g., Meenderink *et al.*, 2005)]. Often the emission type at narrow primary ratio appears mixed, with a shallow slope with superimposed variations. Many studies have been devoted to unmixing these components (e.g., Kalluri and Shera, 2001). By locally indenting the cochlear partition using our micro-pressure-sensor, the local cochlear nonlinearity was reduced, and stiffness perhaps reduced as noted above. The induced changes in cochlear mechanics led to the apparent decrease in the generator-type emission at frequencies somewhat less than the BF of the damaged region and relative increase in the reflector-type at frequencies in the vicinity of the damaged location's BF [Figs. 6(B), 6(D), and 7(B′) and 7(D′)]. Thus, in the BF frequency region, post-damage the phase was steep and the reflector-type emission appeared to be dominant. We evaluate this damage-induced phase change in Fig. 10, which shows a group of DPOAE phase results under pre- (solid lines) and post-damage conditions (dotted lines). Results from six animals are shown. A representative curve showing twice the forward traveling wave phase at the 20 kHz BF from one animal is also shown (gray line). The emission phase post-damage is quite similar to twice the forward traveling wave phase, confirming the notion that the rapidly varying emission phase is representative of the traveling wave in the cochlea, and can be used to estimate traveling wave delays.

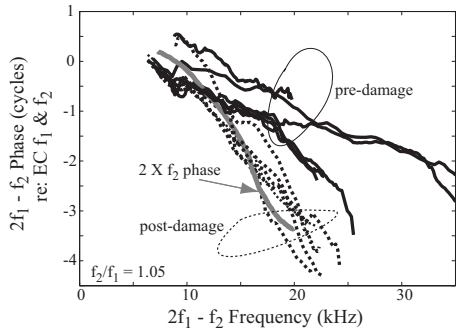


FIG. 10. DPOAE pre- and post-damage phases. Data from experiments wg91, wg92, wg93, wg95, wg122, and wg131. The 20 kHz region was damaged either by locally indenting the BM or by the invasive intracochlear measurement. Solid and dashed lines represent the phase under healthy and locally damaged conditions, respectively. The DPOAE shown is at frequency $2f_1-f_2$ with $L_1=L_2=80$ or 90 dB and $f_2/f_1=1.05$. The gray line shows an example of two times forward f_2 phase with BF ~20 kHz.

V. SUMMARY

Locally indenting the cochlear partition raised CAP thresholds at frequencies within the targeted region and reduced local cochlear nonlinearity, including local DPs, and appeared to reduce the stiffness of the cochlear partition slightly. DPOAE changes following the local damage corresponded to the damaged region, with reductions at some frequencies and enhancements at others. The changes could be interpreted within the two-source framework of cochlear emissions, and indicated an increase in reflector- and a decrease in generator-type emissions. In addition, the results confirmed that for the low frequency sidebands, e.g., $2f_1-f_2$, the generator-type dominates the DPOAE at relatively wide primary ratio (1.25) and both the generator- and reflector-type emissions contribute substantially at narrow primary ratio (1.05).

ACKNOWLEDGMENTS

We gratefully acknowledge the efforts of Glenis R. Long, Glen Martin, and an anonymous reviewer, who provided valuable comments on the manuscript. This work was funded by Grant No. 5R03DC009080 from the NIH/NIDCD.

Avan, P., Bonfils, P., Gilain, L., and Mom, T. (**2003**). "Physiopathological significance of distortion-product otoacoustic emissions at 2 f1-f2 produced by high- versus low-level stimuli," J. Acoust. Soc. Am. **113**, 430–441.

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. **509**, 277–288.

Cooper, N. P. (**2003**). "Compression in the peripheral auditory system," in *Compression From Cochlear to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea," J. Neurophysiol. **78**, 261–270.

Dallos, P. (**1992**). "The active cochlea," J. Neurosci. **12**, 4575–4585.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen-Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1266.

de Boer, E., Zheng, J., Porsov, E., and Nuttall, A. L. (**2008**). "Inverted direction of wave propagation (IDWP) in the cochlea," J. Acoust. Soc. Am. **123**, 1513–1521.

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.

Dong, W., and Olson, E. S. (**2005a**). "Tuning and travel of two tone distor-

tion in intracochlear pressure," in *Auditory Mechanisms: Processes and Models*, *The Ninth International Symposium*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Portland, OR), pp. 56–62.

Dong, W., and Olson, E. S. (**2005b**). "Two-tone distortion in intracochlear pressure," J. Acoust. Soc. Am. **117**, 2999–3015.

Dong, W., and Olson, E. S. (**2006**). "Middle ear forward and reverse transmission in gerbil," J. Neurophysiol. **95**, 2951–2961.

Dong, W., and Olson, E. S. (**2008**). "Supporting evidence for reverse cochlear traveling waves," J. Acoust. Soc. Am. **123**, 222–240.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). "Nonlinear interactions that could explain distortion product interference response areas," J. Acoust. Soc. Am. **108**, 1786–1802.

Harding, G. W., and Bohne, B. A. (**2004**). "Temporary DPOAE level shifts, ABR threshold shifts and histopathological damage following below-critical-level noise exposures," Hear. Res. **196**, 94–108.

Harding, G. W., Bohne, B. A., and Ahmad, M. (**2002**). "DPOAE level shifts and ABR threshold shifts compared to detailed analysis of histopathological damage from noise," Hear. Res. **174**, 158–171.

Henley, C. M., Weatherly, R. A., Martin, G. K., and Lonsbury-Martin, B. (**1996**). "Sensitive developmental periods for kanamycin ototoxic effects on distortion-product otoacoustic emissions," Hear. Res. **98**, 93–103.

Jock, B. M., Hamernik, R. P., Aldrich, L. G., Ahroon, W. A., Petriello, K. L., and Johnson, A. R. (**1996**). "Evoked-potential thresholds and cubic distortion product otoacoustic emissions in the chinchilla following carboplatin treatment and noise exposure," Hear. Res. **96**, 179–190.

Kakigi, A., Hirakawa, H., Harel, N., Mount, R. J., and Harrison, R. V. (**1998a**). "Basal cochlear lesions result in increased amplitude of otoacoustic emissions," Audiol. Neuro-Otol. **3**, 361–372.

Kakigi, A., Hirakawa, H., Harel, N., Mount, R. J., and Harrison, R. V. (**1998b**). "Comparison of distortion-product and transient evoked otoacoustic emissions with ABR threshold shift in chinchillas with ototoxic damage," Auris Nasus Larynx **25**, 223–232.

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.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

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 f_1,f_2 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.

Kolston, P. J. (**2000**). "The importance of phase data and model dimensionality to cochlear mechanics," Hear. Res. **145**, 25–36.

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.

Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613–1626.

Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). "The clinical utility of distortion-product otoacoustic emissions," Ear Hear. **11**, 144–154.

Martin, G. K., Stagner, B. B., Fahey, P. F., and Lonsbury-Martin, B. L. (**2008**). "Steep and shallow phase gradient distortion product otoacoustic emissions arising basal to the primary tones," J. Acoust. Soc. Am. **125**, EL85–EL92.

Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). "Suppression and enhancement of distortion-product otoacoustic emissions by interference tones above f_2. I. Basic findings in rabbits," Hear. Res. **136**, 105–123.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f2/f1," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Meenderink, S. W., Narins, P. M., and van Dijk, P. (**2005**). "Detailed f_1, f_2 area study of distortion product otoacoustic emissions in the frog," J. Assoc. Res. Otolaryngol. **6**, 37–47.

Mills, D. M. (**2002**). "Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response," J. Acoust. Soc. Am. **112**, 1545–1560.

Mills, D. M. (**2003**). "Differential responses to acoustic damage and furosemide in auditory brainstem and otoacoustic emission measures," J. Acoust. Soc. Am. **113**, 914–924.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.

Olson, E. S. (**1999**). "Direct measurement of intra-cochlear pressure waves," Nature (London) **402**, 526–529.

Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.

Olson, E. S. (**2004**). "Harmonic distortion in intracochlear pressure and its analysis to explore the cochlear amplifier," J. Acoust. Soc. Am. **115**, 1230–1241.

Patuzzi, R. B. (**1986**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Rhode, W. S. (**2007**). "Distortion product otoacoustic emissions and basilar membrane vibration in the 6–9 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **122**, 2725–2737.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). "Two-tone distortion on the basilar membrane of the chinchilla cochlea," J. Neurophysiol. **77**, 2385–2399.

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.

Shera, C. A., and Guinan, J. J., Jr. (**2007a**). "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," J. Acoust. Soc. Am. **121**, 1003–1016.

Shera, C. A., and Guinan, J. J., Jr. (**2007b**). "Mechanisms of mammalian otoacoustic emission," in *Active Processes and Otoacoustic Emissions in Hearing*, edited by G. A. Manley, R. R. Fay, and A. N. Popper (Springer, New York), pp. 305–342.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2008**). "Testing coherent reflection in chinchilla: Auditory-nerve responses predict stimulus-frequency emissions," J. Acoust. Soc. Am. **123**, 3851.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.

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.

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., and Nostrant, A. (**1996**). "Selective inner hair cell loss does not alter distortion product otoacoustic emissions," Hear. Res. **96**, 71–82.

Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.

Zhang, X., and Mountain, D. C. (**2008**). "Distortion product emissions: Where do they come from?," in *Concepts and Challenges in the Biophysics of Hearing*, *The Tenth International Workshop on the Mechanics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Keele, Staffordshire, UK), pp. 48–54.

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.

On the significance of phase in the short term Fourier spectrum for speech intelligibility

Michiko Kazama, Satoru Gotoh, and Mikio Tohyama
Waseda University, 161 Nishi-waseda, Shinjuku-ku, Tokyo 169–8050, Japan

Tammo Houtgast
VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 18 November 2008; revised 15 October 2009; accepted 29 December 2009)

This paper investigates the significance of the magnitude or the phase in the short term Fourier spectrum for speech intelligibility as a function of the time-window length. For a wide range of window lengths (1/16–2048 ms), two hybrid signals were obtained by a cross-wise combination of the magnitude and phase spectra of speech and white noise. Speech intelligibility data showed the significance of the phase spectrum for longer windows (>256 ms) and for very short windows (<4 ms), and that of the magnitude spectrum for medium-range window lengths. The hybrid signals used in the intelligibility test were analyzed in terms of the preservation of the original narrow-band speech envelopes. Correlations between the narrow-band envelopes of the original speech and the hybrid signals show a similar pattern as a function of window length. This result illustrates the importance of the preservation of narrow-band envelopes for speech intelligibility. The observed significance of the phase spectrum in recovering the narrow-band envelopes for the long term windows and for the very short term windows is discussed.
 [DOI: 10.1121/1.3294554]

PACS number(s): 43.66.Nm, 43.72.Ja [MW] Pages: 1432–1439

I. INTRODUCTION

The research question of this paper is the significance of the magnitude and the phase in the short term Fourier spectrum for speech intelligibility. The magnitude spectrum is considered important in almost all types of applications of speech processing, while phase has received less attention. Speech signals are commonly analyzed using short-time Fourier transforms (STFTs), and their characteristics are conventionally represented by the magnitude spectrum for speech analysis and/or synthesis.[1] The difference between phonemes is reflected in the structure of the magnitude spectra, and power (or magnitude) spectrum subtraction is commonly used for noise reduction.[2,3]

From a point of view of audio engineering type of applications, it is well known that the phase spectrum of sound is important for rendering (or removing) room-reverberation effects from a reverberation-free (or reverberant) signal. It seems that human listeners are able to appropriately detect phase changes in longer signal segments than those commonly used for speech analysis.[4–8] Schroeder and Strube[5] and Traumueller and Schouten[9] reported that vowels could be synthesized using phase information, and Oppenheim and Lim[10] found that when a speech signal was of sufficient length, speech intelligibility was lost in Fourier-transform magnitude-only reconstruction but not in phase-only reconstruction. However, it is still unclear how effectively the phase information is used for the synthesis of intelligible speech, and formal listening tests were not performed.

On the other hand, Liu *et al.*[11] intensively investigated the effect of the phase on intervocalic stop consonant perception for VCV speech signals. It was shown that the perception of intervocalic stop consonants varies from magnitude dominance to phase dominance as the Fourier-analysis window size increases across the cross-over point between 192 and 256 ms. An effect of phase on perception was also observed for shorter time windows, in the range of 10 to 30 ms. The present study builds on and expands the Liu *et al.* study with respect to two main issues: (1) it investigates the significance of magnitude versus phase in the short term Fourier spectrum for *sentence intelligibility,* rather than for the perception of intervocalic consonants, and (2) it covers a wider range of time-window lengths, with a window covering the complete sentence as the upper limit, down to the lower limit of a single-sample window length.

It is generally believed that speech intelligibility is related to narrow-band envelopes.[12] Drullman[13] found that intelligible speech signals can be synthesized by modulating 24 1/4-octave noise bands covering 100–6400 Hz range, using the temporal speech envelopes obtained in the corresponding 1/4-octave bands. This was the motivation for investigating the role of the phase spectrum in relation to the preservation of narrow-band envelopes. For instance, it will be shown that in case where the window length used in the Fourier transform is substantially larger than the period of the envelope modulation of interest, it is the phase spectrum that carries the information about the temporal envelope, not the magnitude spectrum. It will be shown that the same applies in the case of very short window lengths.

The experimental approach adopted in this paper is similar to the one used by Liu *et al.*,[11] but applied to a spoken sentence and random noise. From these signals, two new signals were created by a cross-wise combination of the magnitude and phase spectra of the speech and noise signals. These two hybrid signals are made for a wide range of win-

 0001-4966/2010/127(3)/1432/8/$25.00

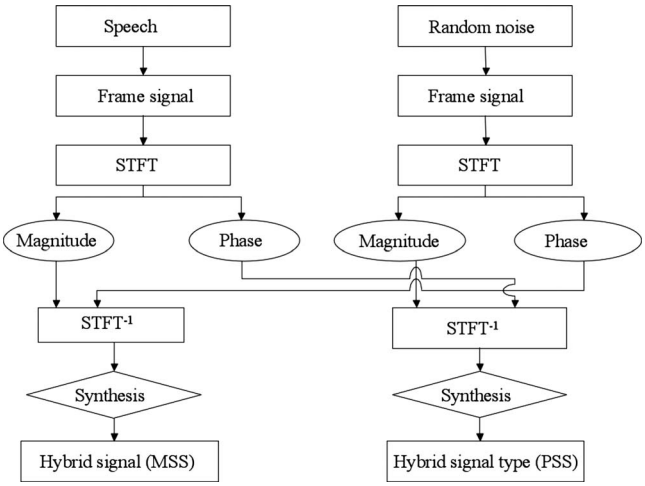


FIG. 1. Method for deriving two types of hybrid signals from speech and random noise using a cross-wise combination of the amplitude and phase spectra in the STFT overlap-add procedure.

dow lengths used in the STFT overlap-add procedure. Sentence intelligibility tests and envelope-correlation studies are performed to investigate the characteristics of the two hybrid signals as a function of window length.

II. LISTENING EXPERIMENT

Synthesized hybrid (magnitude- or phase-only) speech signals were obtained by using female-spoken speech and random-noise samples, as shown in Fig. 1. Sentence intelligibility for the two hybrid signals, as a function of the window length used in the STFT analysis and reconstruction, was estimated using listening tests.

A. Method

1. Test materials and signal processing

The original speech signals consisted of 96 sentences spoken by two female speakers. All of the speech materials were in Japanese and digitized at a sampling rate of 16 kHz, which seems well suited to the sentence intelligibility tests carried out in this study, using a 16-bit A/D converter. Each 1.5-s-long speech phrase had additional silent parts at the start and end, so the total length was 4 s. The speech tokens were simple everyday sentences, with a length of typically six to ten words, e.g., (translated), "This letter cannot be clearly seen from far away." A white-noise signal was produced using MATLAB software.

The speech and random-noise pairs were analyzed using STFT (Fig. 1) where a rectangular-window function was applied to cut the signals into frames. A 50% overlapped windowing was applied (except for the two- or the single-point frames). Two hybrid signals were synthesized by inverse STFT using the magnitude spectrum of the speech (or the noise) and the phase spectrum of the noise (or the speech). The first type will be referred to as magnitude-spectrum speech (MSS) and the second type as phase-spectrum speech (PSS). A triangular window, with a frame length equal to the rectangular window used for the analysis, was applied to each synthesized frame to avoid discontinuities between successive frames. Sixteen frame lengths (1/16–2048 ms) were used, including the limit case of only a single-sample time frame of 1/16 ms. The total set of materials consisted of 192 processed sentences (96×2): six sentences for each of the 16 frame lengths and two types of hybrid signals.

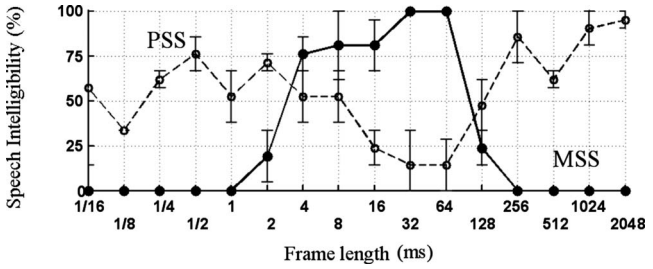


FIG. 2. Sentence intelligibility for PSS and MSS, as a function of the frame length used in the STFT procedure.

2. Subjects and procedure

The listeners were seven men, aged 24–49. They were all native speakers of Japanese. The total set of processed sentences were presented in random order through headphones (AKG-K240) under diotic listening conditions at an individually preferred level. Each subject was asked to write down the sentences as they listened. A sentence was considered intelligible only if the complete sentence was written down correctly.

B. Results

Figure 2 shows the sentence intelligibility scores (with the standard deviation) for each signal type and frame length. The horizontal axis shows the frame length. Each data point is based on six presentations to seven listeners. A score of 100% indicates that all subjects could correctly write down each sentence.

The MSS hybrid signal shows the strongest effect of frame length, ranging from perfectly intelligible for medium-range frames (4–64 ms) to totally unintelligible for long frames (256–2048 ms) and very short frames (1/16–1 ms). The PSS signals show the opposite behavior, though less extreme. For the shorter time frames, the results above suggest that frequency resolution finer than 250 Hz (frame length longer than 4 ms) is needed to get intelligible speech from the spectral magnitude. For the longer time frames, the results suggest that the temporal resolution required to obtain intelligible speech from the magnitude spectrum should be better than about 128 ms, corresponding to a modulation frequency of 8 Hz.

It is interesting to note that, where the magnitude spectrum fails in reproducing intelligible speech, the phase spectrum (partly) takes over this role. For the longer time frames, this corresponds with the earlier observations of Oppenheim and Lim[10] and Liu *et al.*[11] This is consistent with the idea that the temporal properties or signal dynamics represented by the envelopes can be expressed as the very local characteristics of the phase spectrum, such as the group delay. That

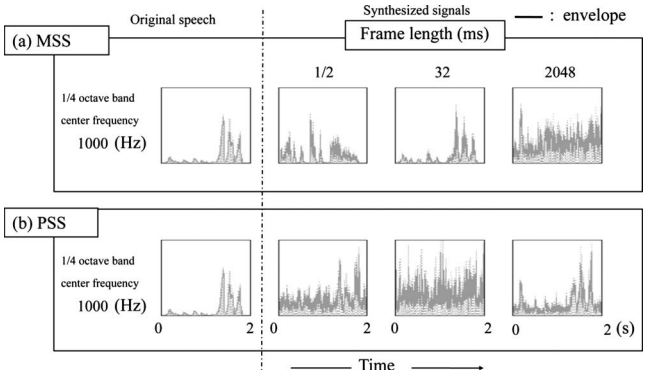


FIG. 3. Samples of squared sub-band waveforms with envelopes for the original speech, and for the MSS and PSS synthesized signals for three frame lengths used in the STFT. The sub-band considered is the 1/4-octave band with 1 kHz center frequency.

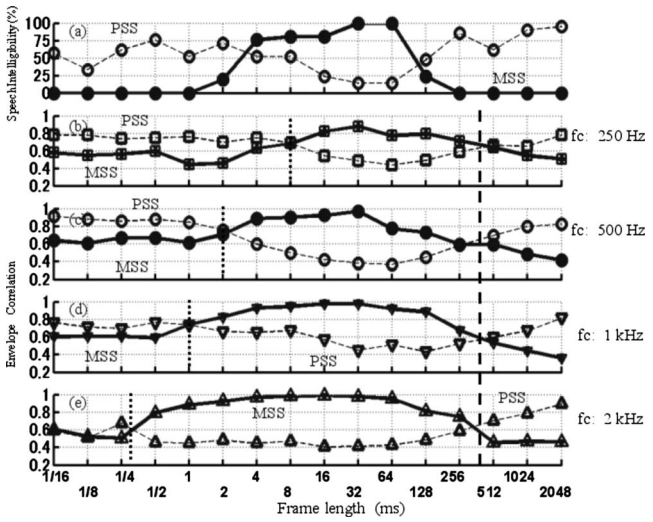


FIG. 4. Sentence intelligibility (a) and examples of envelope-correlation analysis [(b)–(e)] for MSS and PSS. Envelope correlation analysis was made in 1/4-octave bands following Eq. (1) in the text.

is, phase spectra with a fine spectral-resolution (as resulting from long time frames) will allow a partial reconstruction of the narrow-band temporal envelope.

The observed significance of the phase spectrum for the very short time frames is more surprising. This will be discussed later.

III. PRESERVATION OF NARROW-BAND ENVELOPES

As mentioned before, it is generally believed that preservation of intelligibility is related to preservation of narrow-band envelopes.[12–14] In this section, it will be investigated to what extent the narrow-band envelopes are preserved for the two types of hybrid signals. Sub-band signals of 1/4-octave bands between 250 and 4000 Hz were derived by applying a finite impulse response (FIR) filter bank (fourth-order Butterworth filter).[13] The envelope in each frequency band was defined by a Hilbert transform.

A. Observation of synthesized signals and their envelopes

Examples of the envelopes of the original and the hybrid speech signals of MSS and PSS are shown in Fig. 3 for one of the sentences in the stimulus set. This example takes one frequency band (1 kHz 1/4-octave band) and three choices of the time window (1/2, 32, and 2048 ms), motivated by the intelligibility data in Fig. 2. The narrow-band envelope of the MSS illustrated in Fig. 3(a) resembles the original envelope, only for the frame length of 32 ms. The envelope samples of PSS [Fig. 3(b)] show the opposite behavior: resemblance with the original envelope is seen only for the very short and long time frames. The observed qualitative agreement between envelope preservation and the intelligibility data motivated a more detailed study of the two types of hybrid signals in terms of narrow-band envelope correlations as a function of window length.

B. Narrow-band envelope-correlation analysis

The narrow-band envelope-correlation analysis is performed between the original and synthesized speech materials. The nature of the narrow-band temporal envelopes of the signals was evaluated by determining the correlation coefficients between the original and hybrid-signal envelopes. The correlation was calculated for every 1/4-octave band. The correlation coefficient of the ith frequency band for a sentence l is defined as

$$\rho_i(l) = \frac{\overline{\hat{e}_{oi}(l,n)\hat{e}_{si}(l,n)}}{\sqrt{\hat{E}_{oi}(l)\hat{E}_{si}(l)}}, \tag{1}$$

where

$$\hat{e}_{oi}(l,n) = e_{oi}(l,n) - \overline{e_{oi}(l,n)}, \tag{2}$$

$$\hat{e}_{si}(l,n) = e_{si}(l,n) - \overline{e_{si}(l,n)}, \tag{3}$$

$$\hat{E}_{oi}(l) = \overline{\hat{e}_{oi}(l,n)^2}, \tag{4}$$

$$\hat{E}_{si}(l) = \overline{\hat{e}_{oi}(l,n)^2}, \tag{5}$$

and $e_{oi}(l,n)$ and $e_{si}(l,n)$ denote the squared envelopes of the original and synthesized speech signals in the ith band for the sentence l, and the over-line shows taking an appropriate time average. For each frame length and for each of the two hybrid-signal types, the average was taken of the correlation coefficients for the six sentences used for that condition.

Figures 4(b)–4(e) present examples of the correlation coefficients between the narrow-band envelopes of the hybrid signals and the original speech for each of four 1/4-octave bands. Figure 4(a) is just a replication of the intelligibility test results in Fig. 2. The pattern of the correlation coefficients, as a function of the time-window length, is somewhat frequency-band dependent, but the complementary nature of the correlation values for MSS and PSS is observed for each frequency band. The intelligibility data and the narrow-band envelope correlations show the same trend with respect to the effect of frame length. This correspondence between the intelligibility data and the narrow-band envelope data confirm that the preservation of the narrow-band temporal envelopes is closely related to speech intelligibility.

The correlation data for MSS and PSS show two cross-over points. The cross-over point at a frame length of about 256 ms is almost independent of the frequency band considered, as can be seen by the vertical broken line through the figures. Since the observed decrease in the correlation for MSS toward long frame lengths reflects the loss of time resolution required for representing the temporal envelope, this crossover point is supposed to be related to the dominant frequency of the envelope modulations. The corresponding cross-over point in the intelligibility data is considerably lower, suggesting that the speech envelope includes slow modulations, which are included in the correlation values, but contribute little to speech intelligibility. This point is addressed in Sec. IV.

The other cross-over point is frequency dependent as can be seen by the vertical dotted lines in each of the figures. The cross-over points happen to correspond roughly with the duration of the period of the center frequency. We cannot provide a firm theoretical basis for this relation. In general terms, however, the frequency dependency of the cross-over point can be understood as a reflection of the limited frequency resolution associated with a short frame length. Given the increase in bandwidth for increasing center frequency f_c of the 1/4-octave bands considered in Fig. 4, a certain loss of frequency resolution (typically equal to the inverse of the frame length in the STFT) will have less effect for higher $f_c's$. Thus, in order to recover 1/4-octave band envelopes from the magnitude spectrum, the frame length used in the STFT should provide an adequate degree of frequency resolution, related to the width of the frequency band considered. Hence, shorter frames are allowed toward higher $f_c's$.

The results so far can be summarized as follows.

(1) *The MSS data are quite understandable*. For longer time frames (>256 ms), the temporal resolution is insufficient to follow the relevant envelope modulations, and for shorter time frames (<4 ms), the frequency resolution becomes insufficient (this appears to depend on the center frequency of a band).
(2) *The PSS data are more surprising*. The envelope is (partly) recovered for windows longer than 256 ms, and also for the very short time frames (which may not be intuitively obvious for many readers).

These observations on the phase dominance for longer and for very short time frames will be studied further by analyzing narrow-band envelope recovery from the phase spectrum only.

C. Recovery of narrow-band envelopes from the phase spectrum

1. Significance of phase spectrum for long window lengths

The importance of the phase spectrum for modulated signals is well illustrated by the difference between an amplitude- and a quasi-frequency-modulated (AM and QFM) sinusoid. It is well known that the phases of the two side-band components determine the temporal envelope: essentially flat in the QFM case and modulated in the AM case. Figures 5(a) and 5(b) show a stationary random noise and a noise modulated by a co-sinusoidal function, respectively. The corresponding magnitudes and phase spectra are shown in Figs. 5(c)–5(f). The envelope-modulation frequency in this example is given by $2(1/N)$, where N denotes the signal length. Here, STFT analysis was applied to the whole signal length.

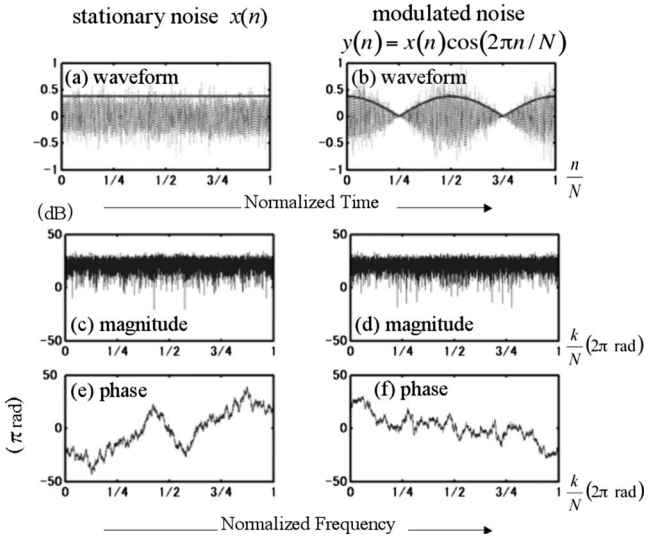


FIG. 5. Examples of stationary random noise (a) and modulated noise (b) with the magnitude [(c) and (d)] and phase [(e) and (f)] spectral characteristics.

Although there are no clear indications of the envelope frequency in the magnitude and phase spectra, the frequency can be observed by applying an auto-correlation analysis to the phase spectrum. This is illustrated in Fig. 6, where the modulation frequency is converted to a real quantity. When the phase difference between components (k_o) and $(k+k_o)$ is given by

$$\Delta\theta(k,k_o) = \theta(k+k_o) - \theta(k_o), \tag{6}$$

then the phase correlation function phc(k) can be obtained by

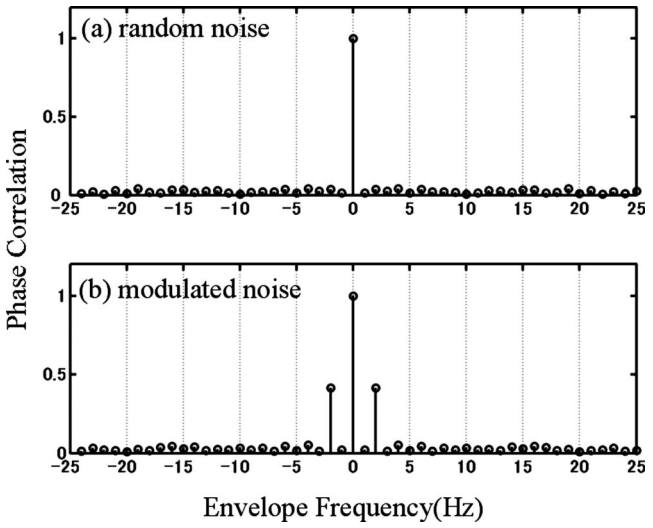


FIG. 6. Phase spectrum auto-correlation analysis for the signals shown in Fig. 5 according to Eqs. (6)–(9) in the text.

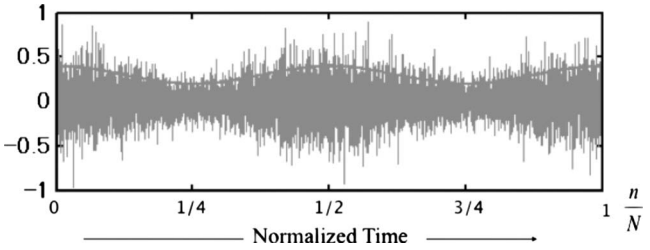


FIG. 7. Reconstruction of the modulated noise of Fig. 5(b), using the corresponding phase spectrum and a random magnitude spectrum.

$$\mathrm{phc}_c(k) = \frac{1}{K}\sum_{k_o=0}^{k_o=K-1} \cos\Delta\theta(k,k_o), \tag{7}$$

$$\mathrm{phc}_s(k) = \frac{1}{K}\sum_{k_o=0}^{k_o=K-1} \sin\Delta\theta(k,k_o), \tag{8}$$

$$\mathrm{phc}(k) = \sqrt{\mathrm{phc}_c(k)^2 + \mathrm{phc}_s(k)^2} \tag{9}$$

in a discrete form. Here, K denotes the number of frequency components of interest. In the figure, the horizontal axis shows the frequency shift, which can be interpreted as the envelope frequency.

Note that Fig. 6(a) shows that the fluctuations seen in the phase spectrum for stationary noise are random. Only the modulated-noise case [Fig. 6(b)] shows that the modulation frequency can be estimated from phase information alone. Figure 7 is an example of a hybrid signal made by substituting random magnitude for the original magnitude spectrum of the modulated noise shown in Fig. 5(b). This illustrates that the original envelope is partly preserved on the basis of the phase spectrum only. The spacing of the frequency components in the phase spectrum resulting from the STFT should be small enough to reflect the envelope frequency in the phase spectrum auto-correlation function. Since this frequency spacing is related to the inverse of the frame length used in the STFT, this implies that for envelope recovery from the phase spectrum, the frame length should be longer than the period of the envelope modulation of interest.

2. Significance of phase spectrum for very short window lengths

Phase dominance for very short frame lengths can be interpreted as the narrow-band envelope recovery from the zero-crossings of a waveform. As Figs. 4(b)–4(e) indicate, this requires that the frame length is shorter than the period of the center frequency of interest. For the present study, the limit case of a very short analysis window-lengths is a length of 1/16 ms (i.e., the sampling rate), corresponding to a single-point STFT. The result of a single-point STFT is for each sample, its magnitude, and the phase is just the sign of the sample, ±. Thus, the phase information of a single-point STFT keeps the zero crossings of the original signal, if the sample frequency is adequate. This is the same as applying infinite peak clipping to a signal, which also preserves the zero-crossing information while losing all amplitude information. It will be shown that the narrow-band envelopes can be partly recovered from the zero crossings of the original signal.

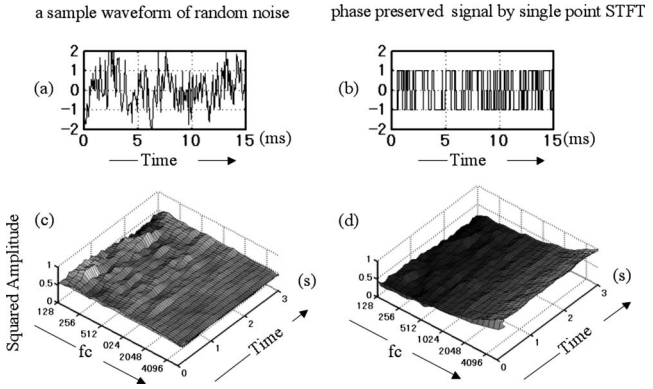


FIG. 8. Representation of stationary noise. Left part: original. Right part: after single-point STFT, with the phase (+ or −) of each sample preserved, and the magnitude set at unity.

Figure 8(a) shows a snap shot of a waveform from a stationary random noise. Figure 8(b) represents the random noise re-synthesized by preserving the phase only for a single point of STFT (just ±) with the magnitude of unity. This resulted in an infinitely peak-clipped version of the original signal. The three dimensional plots in Figs. 8(c) and 8(d) show the temporal change in the short-time sub-band energy of these two signals, indicating that the original spectro-temporal characteristics, especially in the low frequency range, are preserved to some extent in the signal synthesized from the phase information of the single-point STFT. The increase in the energy for the high frequency bands in Fig. 8(d) can be interpreted as a processing noise due to the hard clipping of the waveform shown in Fig. 8(b).

The example above refers to a stationary signal. It illustrates that the shape of the spectral magnitude information is hidden in the pattern of the zero-crossings, particularly for the low frequency bands. Figure 9 is another example for a random noise, but now including a modulated sub-band. The original zero-crossings are preserved in Fig. 9(b) by the single-point STFT. Despite losing the original magnitude information, the temporal envelope of the sub-band can be recovered to some extent from the zero-crossing information, as shown in Fig. 9(d). In other words, the narrow-band en-

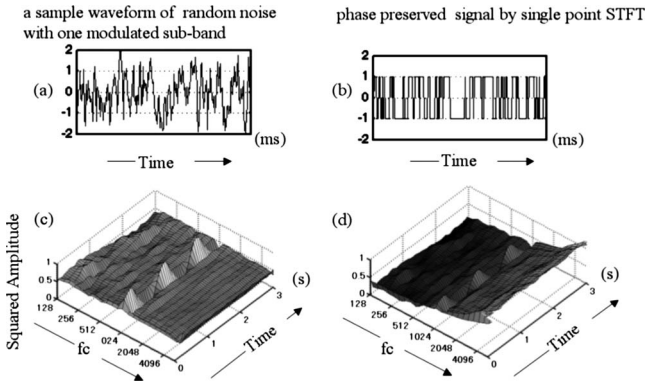


FIG. 9. A sample waveform of random noise with a modulated sub-band. Left part: original. Right part: similar to Fig. 8.

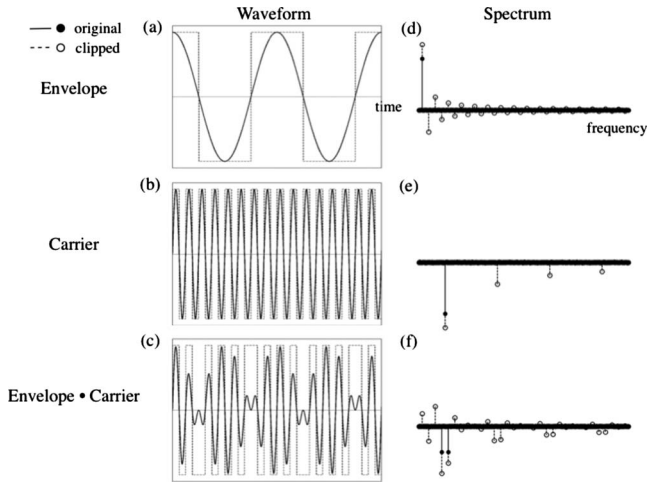


FIG. 10. Spectrum of the infinitely clipped version of a modulated sinusoidal signal.

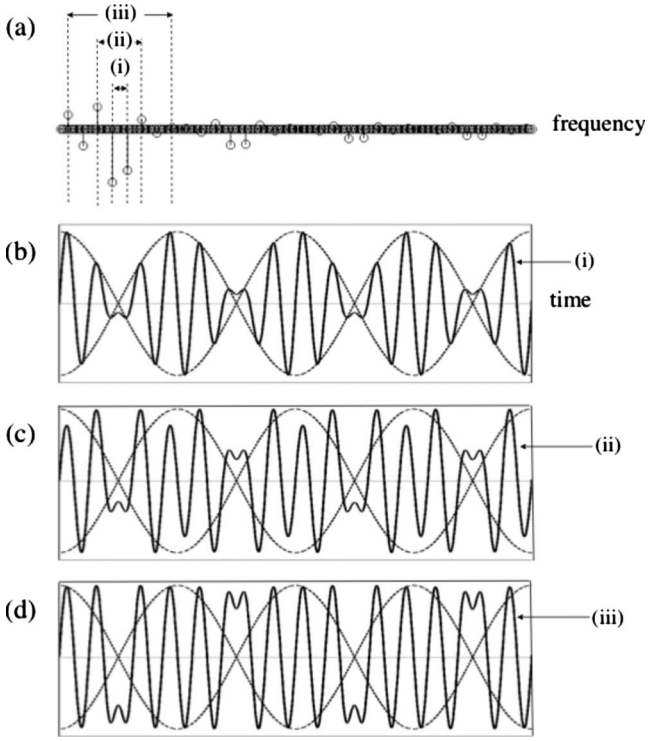


FIG. 11. Sinusoidal envelope recovery from the clipped wave as shown in Fig. 10, after applying sub-band filtering with increasing bandwidth, indicated by (i), (ii), and (iii).

velope can be partly recovered from the fine-structure (zero crossings) of the modulated noise samples after sub-band analysis.

This example may explain the significance of the phase for the very short time frames, as observed in Fig. 4. However, Fig. 4 also indicates that for the very short time frames the envelopes for the higher frequency bands are not well recovered from the phase-only information. This may be related to the fact that, given the typical shape of the power spectrum of speech, the higher bands represent only a small fraction of the total power, and consequently the modulation properties of these higher bands may only be marginally represented in the zero-crossing statistics of the over-all signal. Another possibility is that envelope reconstruction from phase is disrupted by the processing noise that yields higher energy at higher frequencies, as shown in Figs. 8(c), 8(d), 9(c), and 9(d).

Another example of envelope recovery from zero-crossing information is provided in Figs. 10 and 11. It concerns a modulated sinusoidal waveform, as shown in Figs. 10(a)–10(c). Spectral records for the envelope [Fig. 10(d)], its carrier [Fig. 10(e)], and the modulated signal [Fig. 10(f)] are represented by the line-spectral characteristics. Here, the solid lines and solid circles show the original ones, while the dotted lines and open circles indicate the infinitely clipped case. The spectral structure of the modulated signal can be expressed as the convolution of the spectral sequences for the envelope and the carrier. The convolution is performed periodically, because this numerical sample is composed of a discrete sequence. If only the zero-crossing property is preserved with the magnitude of unity (discarding the envelope of the modulated sinusoid), the convolved spectral-structure is expanded, including its higher harmonics. Although those higher harmonics are not contained in the original modulated signal, the modulation property, such as the temporal envelope, can be recovered by applying appropriate filtering, as shown in Fig. 11.

Figure 11(a) is close-up of Fig. 10(f). If we take a bandwidth denoted by (i) in the figure, representing sub-band analysis, then we get the waveform shown in Fig. 11(b). Here, the broken line represents the original envelope shown in Fig. 10(a). However, if we increase the bandwidth according to the examples denoted by (ii) or (iii) in Figs. 11(c) and 11(d), the original envelope is no longer recovered. This illustrates that the original envelope can be recovered from zero-crossing information when applying sub-band filtering, provided that the bandwidth is adapted to the modulation frequency of interest. Since in our analysis, higher frequencies are associated with broader absolute bandwidth, this may be a reason why envelope recovery from phase for very short windows is poorer at high frequencies (Fig. 4), and why the processing noise increases toward high frequencies [Figs. 8(d) and 9(d)].

In principle, characteristics of speech waveforms can be understood as a mixture of a random noise-like feature or a periodic structure. Thus, the two simplified examples presented above represent two extreme cases of signals with speech-like characteristics. For both some form of sub-band-envelope recovery from the zero-crossing information only was demonstrated.

IV. DISCUSSION

The results of the listening experiment, as presented in Fig. 2, provide the key data of the present study. We will first consider the strong effect shown in the MSS data, in particular, the decrease to 0% intelligibility for the long and the short frame lengths when using the magnitude spectra only.

In interpreting the loss of intelligibility of MSS speech for time windows of over about 250 ms, it was assumed that this reflects the loss of temporal resolution required for following the essential speech-envelope modulations. This

would suggest that the envelope modulations above 4 Hz are indispensable for maintaining intelligible speech. A related study on the effect of low-pass filtering narrow-band envelopes[15] indicated that only for a 2-Hz (or lower) low-pass cut-off frequency is sentence intelligibility severely reduced. The difference may be understood by realizing that, in this previous study, the envelope filtering was the only "distortion" applied to the speech, while in the present study many additional "distortions" are introduced in the MSS condition, as a result of disregarding the phase spectrum.

The loss of intelligibility of MSS speech for time windows of 1 ms or less was interpreted as reflecting the very limited frequency resolution associated with such brief time windows. This would imply that a frequency resolution of worse than about 1000 Hz makes speech unintelligible. In related studies, e.g., on the effect of spectral smearing on speech intelligibility,[16] or on the minimum number of bands required to produce intelligible speech,[17] the operations are performed on a logarithmic frequency scale, complicating the comparison with the present study (window-associated loss of spectral-resolution is constant over frequency). Still, the 1000-Hz limit suggested by the present data does not disagree with the findings of these other studies.

For a frame length in the range of 4–64 ms, the magnitude spectrum carries the essential information for speech intelligibility. Thus, for the commonly observed approach in speech processing, i.e., a spectral analysis with a window length of a few tens of ms, the use of the power spectrum does maintain the essential cues for speech intelligibility. Following the traditional view of the peripheral auditory system as a set of band pass filters, the auditory temporal window in the mid-frequency range amounts to typically 10 ms.[18] Figure 4 indicates that for this window length, both the speech intelligibility and the narrow-band envelope preservation are dominated by the magnitude spectrum, while the phase only plays a minor role. This implies that the envelopes of the auditory filter outputs carry the intelligibility-relevant information.

The main goal of the present study was to investigate the relative importance of magnitude versus phase in the short term Fourier-spectrum approach in speech analysis and synthesis, given that most studies concentrate on the magnitude or power spectrum only. The experimental signal manipulations used in this study (resulting in the hybrid signals) will have consequences for the preservation of the envelope and the fine structure in narrowband (auditory) filters. In this respect, there is a relation between this study and research on the relative importance for speech intelligibility of envelope cues and fine-structure (phase) cues at the auditory filter outputs. The consensus on this issue (see, among others, Refs. 14, 13, and 19) is that preservation of the envelopes at the filter outputs is the main factor for speech intelligibility, while the phase or the fine structure is of secondary importance. This firm distinction is somewhat complicated by studies, indicating that envelope and phase information at the filter outputs are not independent,[20] and that temporal-envelope cues can be recovered from the speech fine structure.[21] A detailed study on the effect of additive noise on speech intelligibility, quantifying the relative importance of disrupting the narrow-band envelopes or the fine structure, confirmed the importance of the envelopes.[22] However, besides the main role of envelope cues, that study also showed some reduction in intelligibility after disrupting the fine structure cues. This is in line with various other studies, indicating that temporal fine-structure cues do play a role in speech intelligibility, especially in case of complex (i.e., non-stationary) maskers.[23–25]

The relevance of narrow-band envelopes for speech intelligibility motivated the use of this concept (i.e., the degree of preservation of narrow-band envelopes) for interpreting the present intelligibility data on the relative importance of the magnitude or phase spectrum. However, it should be realized that, besides the loss of envelope correlation for the pre- and post-processed speech, other types of distortions are introduced as well. For instance, it has been shown that a loss of cross-spectral modulation phase coherence may reduce speech intelligibility.[26] It is very probable that the observed loss of correlation between the pre- and post-processed narrow-band envelopes is associated with a loss of cross-spectral modulation phase coherence. Also, as mentioned before, the processing will affect the narrow-band fine structure. Consequently, part of the observed relation in the present study between loss of intelligibility and loss of band-envelope correlation may well be caused by the associated effects of loss of cross-spectral modulation phase coherence or loss of fine structure. The present study does not allow to specify this any further.

V. CONCLUSION

Speech was subjected to Fourier analysis and synthesis, using the overlap-add procedure with window lengths ranging from 1/16 to 2048 ms. Experiments were performed on the intelligibility of the speech for two conditions applied for the synthesis: (a) MSS-mode, using the speech-magnitude spectra with randomized phase spectra, and (b) PSS-mode, using the speech-phase spectra with randomized magnitude spectra. Besides the intelligibility measurements, the signals were subjected to an analysis of the correlation between the narrow-band envelopes, before and after the MSS or the PSS synthesis mode. The main findings were as follows.

(1) Using the MSS synthesis mode (magnitude spectra only), intelligible speech was obtained only for frame lengths of 4–64 ms.
(2) When using the PSS synthesis mode (phase spectra only), reasonably intelligible speech was obtained for frames longer than 128 ms or shorter than 4 ms.
(3) Thus, the two curves of intelligibility, as a function of frame length for the MSS and PSS synthesis mode, show a complementary character. This means that, for speech intelligibility, for the medium-range windows (4–64 ms), the magnitude spectrum dominates, and for the two extreme regions (<4 and >64 ms), the phase spectrum dominates.
(4) Intelligibility scores and correlation coefficients between the synthesized and original envelopes in 1/4-octave bands showed the same trend with respect to the effect of frame length, although not identical. This qualitative cor-

respondence confirms that the preservation of narrow-band temporal envelopes constitutes an important factor for the preservation of speech intelligibility.

The interpretation of these findings may be summarized as follows.

(a) *Considering the MSS synthesis mode (magnitude spectra only)*. For long time frames, the speech becomes unintelligible due to a loss of time resolution, as the frame length becomes longer than the period of the dominant speech envelope modulations (about 256 ms, a 4-Hz modulation frequency). For frame lengths shorter than 4 ms, the speech becomes unintelligible because the corresponding frequency resolution is insufficient.
(b) *Considering the PSS synthesis mode (phase spectra only)*. For long time frames, it is shown that the phase spectrum contains envelope information, as reflected in phase-spectral auto-correlation analysis. For the very short frames, it is shown that the phase-only synthesized speech essentially keeps the zero-crossing interval information, from which the shape of the original power spectrum can be obtained. It is shown that the temporal envelope of a sub-band can be partly recovered from the zero crossings of the total signal.

The main result of this study is that, besides the dominance of the magnitude spectrum for the middle range of window lengths, there appear to be two regions of phase dominance with respect to intelligibility and preservation of narrow-band envelopes. This phase dominance applies to very short and long time windows.

ACKNOWLEDGMENT

This research was partly supported by the Telecommunications Advancement Research Fellowship, Japan.

[1]M. R. Schroeder, "Computer speech," *Springer Series in Information Sciences* (Springer-Verlag, Berlin Heidelberg, 1999), pp. 63–73.
[2]S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120 (1979).
[3]P. Vary, "Noise suppression by spectral magnitude estimation-mechanism and theoretical limits," Signal Process. **8**, 387–400 (1985).
[4]M. R. Schroeder, "Computer speech," *Springer Series in Information Sciences* (Springer-Verlag, Berlin Heidelberg, 1999), pp. 113–127.
[5]M. Schroeder and H. Strube, "Flat-spectrum speech," J. Acoust. Soc. Am. **79**, 1580–1583 (1986).
[6]M. R. Schroeder, "Models of hearing," Proc. IEEE **63**, 1332–1350 (1975).
[7]H. Pobloth and W. Kleijn, "Squared error as a measure of phase distortion," in Proceedings of the EUROSPEECH (ISCA) (2001), pp. 1973–1976.
[8]R. Plomp and H. Steeneken, "Effect of phase on the timbre of complex tones," J. Acoust. Soc. Am. **46**, 409–421 (1969).
[9]A. Traumueller and M. Schouten, *The Psychophysics of Speech Perception* (Kluwer, Dordrecht, 1987).
[10]A. Oppenheim and J. Lim, "The importance of phase in signals," Proc. IEEE **69**, 529–541 (1981).
[11]L. Liu, J. He, and G. Palm, "Effects of phase on the perception of intervocalic stop consonants," Speech Commun. **22**, 403–417 (1997).
[12]T. Houtgast, H. Steeneken, and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," Acustica **46**, 60–72 (1980).
[13]R. Drullman, "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592 (1995).
[14]R. Shannon, F. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," Science **270**, 303–304 (1995).
[15]R. Drullman, J. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064 (1994).
[16]M. ter Keurs, J. Festen, and R. Plomp, "Effect of spectral envelope smearing on speech reception II," J. Acoust. Soc. Am. **93**, 1547–1552 (1993).
[17]L. Friesen, R. Shannon, D. Baskent, and X. Wang, "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 (2001).
[18]C. Plack and B. Moore, "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187 (1990).
[19]Z. Smith, B. Delgutte, and A. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90 (2002).
[20]O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640 (2001).
[21]G. Gilbert and C. Lorenzi, "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444 (2006).
[22]F. Dubbelboer and T. Houtgast, "A detailed study on the effects of noise on speech intelligibility," J. Acoust. Soc. Am. **122**, 2865–2871 (2007).
[23]D. Gnansia, V. Péan, B. Meyer, and C. Lorenzi, "Effects of spectral smearing and temporal fine structure degradation on speech masking release," J. Acoust. Soc. Am. **125**, 4023–4033 (2009).
[24]C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, and B. Moore, "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869 (2006).
[25]S. Sheft, M. Ardoint, and C. Lorenzi, "Speech identification based on temporal fine structure cues," J. Acoust. Soc. Am. **124**, 562–575 (2008).
[26]S. Greenberg and T. Arai, "The relation between speech intelligibility and the complex modulation spectrum," in Proceedings of the EUROSPEECH (ISCA) (2001), pp. 473–476.

The acoustical bright spot and mislocalization of tones by human listeners

Eric J. Macaulay and William M. Hartmann[a)]
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

Brad Rakerd
Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824

(Received 11 September 2009; revised 27 December 2009; accepted 29 December 2009)

Listeners attempted to localize 1500-Hz sine tones presented in free field from a loudspeaker array, spanning azimuths from 0° (straight ahead) to 90° (extreme right). During this task, the tone levels and phases were measured in the listeners' ear canals. Because of the acoustical bright spot, measured interaural level differences (ILD) were non-monotonic functions of azimuth with a maximum near 55°. In a source-identification task, listeners' localization decisions closely tracked the non-monotonic ILD, and thus became inaccurate at large azimuths. When listeners received training and feedback, their accuracy improved only slightly. In an azimuth-discrimination task, listeners decided whether a first sound was to the left or to the right of a second. The discrimination results also reflected the confusion caused by the non-monotonic ILD, and they could be predicted approximately by a listener's identification results. When the sine tones were amplitude modulated or replaced by narrow bands of noise, interaural time difference (ITD) cues greatly reduced the confusion for most listeners, but not for all. Recognizing the important role of the bright spot requires a reevaluation of the transition between the low-frequency region for localization (mainly ITD) and the high-frequency region (mainly ILD). © *2010 Acoustical Society of America.*
[DOI: 10.1121/1.3294654]

PACS number(s): 43.66.Pn, 43.66.Qp [JCM] Pages: 1440–1449

I. INTRODUCTION

Sine tones with slow onsets or with masked onsets are localized on the basis of ongoing interaural level differences (ILDs) and interaural time differences (ITDs). The ILDs are mainly useful cues at high frequencies, where the head casts a significant shadow on the ear further from the source of sound.

The ITDs appear to provide the dominant localization cues for sine tones in free field at low frequencies (Wightman and Kistler, 1992; Hartmann and Wittenberg, 1996). However, as the frequency of the tone increases toward 1000 Hz and beyond, the ITD becomes an unreliable cue because the corresponding interaural phase difference (IPD) is close to 180°, causing the perceived direction of the ITD cue to be ambiguous. For a frequency of 1000 Hz, an IPD of 180° occurs for an azimuth near 41°, and as the azimuth increases to 50°, the IPD becomes −150°, i.e., reversed in sign.[1] Perhaps because of this phase problem, the human nervous system becomes rapidly insensitive to ITDs in a steady sine tone as the tone frequency increases. Little or no ITD sensitivity is found at 1300 Hz and higher (Zwislocki and Feldman, 1956). There is a strong evolutionary advantage for such insensitivity because if the IPD sign is reversed, the ITD cue indicates a source direction that is opposite to the true direction. Better to have no cue at all than to have such a misleading cue.

The loss of the ITD cue is not the only interaural change that occurs as the frequency of a tone increases beyond 1000 Hz. A second change is that the interaural level difference becomes a seriously non-monotonic function of azimuth because of the peculiar physics of wave diffraction around the head. Figure 1 shows the calculated ILD vs azimuth for six different frequencies. The ILD is a non-monotonic function of azimuth for all six. Theoretically, the ILD function shows non-monotonic behavior for all frequencies (even 500 Hz in Fig. 1), but that behavior becomes perceptually important only for 1000 Hz and higher. For a 1500-Hz tone, the peak ILD (at an azimuth of 50°) is larger than 8 dB. This can be compared to an ILD of less than 4 dB at 90°. The functions in Fig. 1 were calculated from a spherical-head diffraction model (Rschevkin, 1963; Kuhn, 1977) with antipodal ears (±90° from the forward direction). If the ears are further back on the head, the peak moves to smaller azimuths. For instance, for 1500 Hz, the peak moves from 50° to 41° as the ear angle increases from 90° to 100°.

The origin of the non-monotonic ILD is the acoustical "bright spot." According to Fresnel's theory of diffraction, the intensity of a wave diffracted by a sphere is anomalously large at a location on the sphere directly opposite the direction of wave incidence. The bright spot, often called Poisson's bright spot or Arago's bright spot, was an important effect in demonstrating the wave nature of light (Hecht,

a)Author to whom correspondence should be addressed. Electronic mail: hartmann@pa.msu.edu

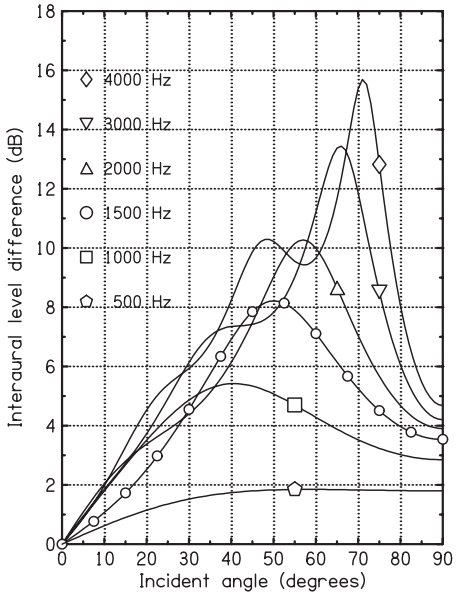


FIG. 1. The interaural level difference plotted as a function of source azimuth (Incident angle) for tones of six different frequencies, as computed in the spherical-head model with antipodal ears. All six curves have a peak, but it is hard to see for 500 Hz.

2002; Kelly *et al.*, 2009).[2] The acoustical effect is shown in panel (b) of Fig. 2 for a 1500-Hz tone. The solid lines show spherical-head calculations. The symbols show measurements made with a Knowles Electronics Manikin for Acoustics Research (KEMAR) (Burkhard and Sachs, 1975). Although the level in the ear near the source tends to grow with increasing azimuth, the level in the ear far from the source first decreases, then increases. The maximum far-ear level at 90° is the bright spot itself, but the approach to the bright spot is seen in the far ear for all angles greater than 50° or 60°. The difference between the near-ear and far-ear levels is the ILD, as shown in Fig. 2(c). Because of the bright spot, the ILD is a non-monotonic function of azimuth. As Kuhn (1987) observed, significant non-monotonic behavior is characteristic of the ILD throughout the high-frequency region.

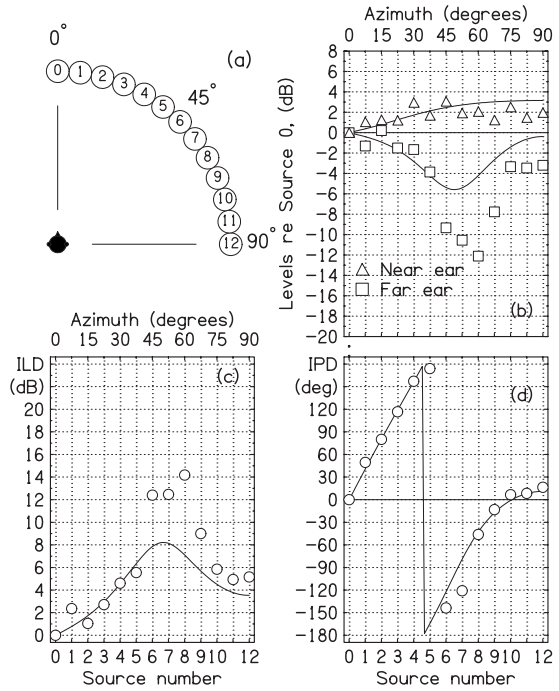


FIG. 2. Spherical-head computations (solid lines) and recordings in KEMAR ears (symbols). The sizes of the symbols indicate about 4 standard deviations. (a) The arrangement of loudspeakers and listener in the anechoic room. (b) The levels in near and far ears. Levels are plotted as functions of source azimuth for a 1500-Hz tone. They are referenced to the level at zero azimuth. (c) The ILD is the difference between near and far curves from part (b). The "near" ear is the right ear, and the "far" ear is the left ear. (d) Interaural phase difference.

The goal of the experiments reported in this article was to determine the perceptual consequences of the bright spot for large azimuths and relatively high frequency. Specifically, we hypothesized that at a frequency of 1500 Hz, the ITD would not be a useful cue and listeners could only rely on the ILD to encode the azimuth. Therefore, we expected that localization judgments would be dramatically misled by the non-monotonic character of the ILD. It was not completely obvious that this hypothesis would be supported experimentally. First, the theoretical curves and data in Fig. 2 might not apply to real listeners. Second, if they do apply, then perhaps listeners recognize the ambiguity of the ILD cues and have learned to cope with them. For example, Fig. 2(b) shows that there might be an opportunity for listeners to use the level information in the two ears separately to solve the azimuthal problem. As McFadden (1981) pointed out, peripheral interaural cues always require interpretation by higher-level functions.

II. EXPERIMENT 1: IDENTIFICATION

In an identification experiment, listeners were asked to identify the location of a loudspeaker producing a 1500-Hz sine tone. A frequency of 1500 Hz was chosen because of the large range of azimuths where the ILD slope is negative, as shown in Figs. 1 and 2(c). Also, 1500 Hz is low enough that the spherical-head model predicts only a simple peak as a function of azimuth.

A. Method

Listeners were seated near the center of an anechoic room (IAC 107840) with interior dimensions of $3.0 \times 4.3 \times 2.4$ m^3. An L-shaped rod mounted to the back of the chair touched the top of the listener's head to help him keep his head position fixed and facing in the forward direction. At a distance of 112 cm from the listener was an arc of 13 single-driver loudspeakers (Minimus 3.5), with azimuths varying in 7.5° increments over the range from 0° (forward direction) to 90°, as shown by the numbered circles in Fig. 2(a). The loudspeaker cones were at the height of the listener's ears, within a few centimeters. Listeners were familiar with the loudspeaker positions. A small drawing of the speaker array, with positions numbered 0–12 [Fig. 2(a)], was mounted below the front speaker (number 0) to remind listeners of the geometry.

Prior to the experiment, the level of each loudspeaker was measured using an omnidirectional microphone at the listener's position with the chair removed. In the experiments that followed, the levels of the signals sent to loudspeakers

were equalized so that all the loudspeakers produced the same level at the listener's position, ±0.5 dB. Of course, the levels in the listener's ears were ultimately not all the same, but depended on the speaker azimuth and the listener's head. The reason for using constant source level (and not roving) was that preliminary experiments suggested that listeners would be confused by the non-monotonic ILD, and we wanted to give the listeners every possible advantage in performing the task. In principle, equal-level sources might allow the listeners to localize by using the levels in the two ears independently if they are able to do so.

During the experiment, a computer program generated the tone, selected one of the 13 sources at random, and played the tone twice via Tucker-Davis digital-to-analog converters (DD1). The listener's task was to listen to the tones and to report the perceived azimuth verbally by an intercom. The experimenter then entered that choice into the computer for later analysis.

The tone level was 72 dBA at the listener's position. To avoid onset and offset cues, the tone had rise and fall durations of 250 ms. Each of the two tones had a full-on duration of 500 ms, and there was a gap of 1.8 s between the two tones. It was expected that presenting the tone twice from the same loudspeaker would make listener judgments more reliable. The presentations were paced by the listener's responses.

Each experimental run consisted of five passes through the set of loudspeakers. In any pass, each of the 13 speakers was presented once in random order. Thus, there were 65 decisions per run. A typical run lasted about 10 min. After a run, the listener could come out of the anechoic room and rest. There was no trial-by-trial feedback.

B. Physical measurements

While the experiment was in progress, each tone was measured using probe microphones in the listener's two ear canals (Etymotic ER-7c system with matching preamplifier). Prior to conversion to digital form (Tucker-Davis DD1), the microphone signals were given an additional 40 dB of gain. The digital recordings were processed by matched filtering, which led to recorded amplitudes and phases in each ear. The left and right amplitudes and phases were used to compute ILD and IPD values. Because each speaker was chosen for five trials in a run and a tone was presented twice for each trial, there were ten independent measurements of ILD and IPD per run. These were then averaged over two runs (20 measurements) to find the interaural parameters for each speaker. All measured values of IPD were reduced to the range from −180° to +180°.

The ILD and IPD values obtained for source number 0 (directly ahead) were small. Their mean values were subtracted from the mean values of all the other sources so that the ILD and IPD values reported in this article represent the changes in interaural parameters caused by the angular displacement away from the forward direction.

Reliability of the acoustical measurements was tested by using the probe microphones to measure the signals in the ear canals of a KEMAR manikin. The probes were inserted

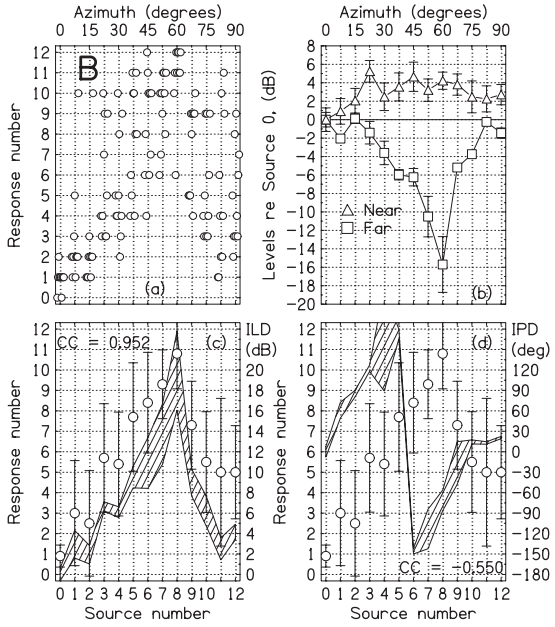


FIG. 3. Identification responses and probe-microphone measurements for listener B. (a) Scatter plot showing all responses. Points have been slightly displaced horizontally for clarity. (b) Measured levels in near (right) and far (left) ears. Error bars are 2 standard deviations in overall length. (c) Listener identification responses on the scale 0–12, with error bars 2 standard deviations in overall length. Perfect performance would correspond to a 45° line. The hatched region shows the ILD. For each source, it is centered on the mean ILD, and the overall height of this region is 2 standard deviations. The ILD values were measured while the response data were being collected. (d) Identification responses as in part (c) plotted together with IPD. The hatched region shows the IPD. For each source, it is centered on the mean IPD, and the overall height is 2 standard deviations. The IPD values were measured while the response data were being collected.

ten different times to simulate different fittings for human listeners, and measurements were made each time. The measured values are shown by the symbols in Fig. 2(b). The standard deviation across the ten different measurements was the size of a symbol or smaller.

C. Listeners

There were five listeners, all male. Listeners E, M, and N were in their twenties. Listener X was 35, and B was 56. All listeners had pure-tone thresholds within 15 dB of normal according to audiometric tests in the range 200–8000 Hz.

D. Results and discussion

In collecting the physical measurements, the experimenters monitored the measured levels and phases to verify correct operation. The standard deviations were examined to make sure that the measurements were not adversely affected by motion of the listener. Although the listeners' head positions were minimally constrained, the data were collected without having to repeat any runs except for one listener.

1. Levels in the ear canals

All data reported here are based on the two runs; a total of ten trials for each loudspeaker. The ear-level data are presented in the (b) parts of Figs. 3–7 for listeners B, E, M, N,

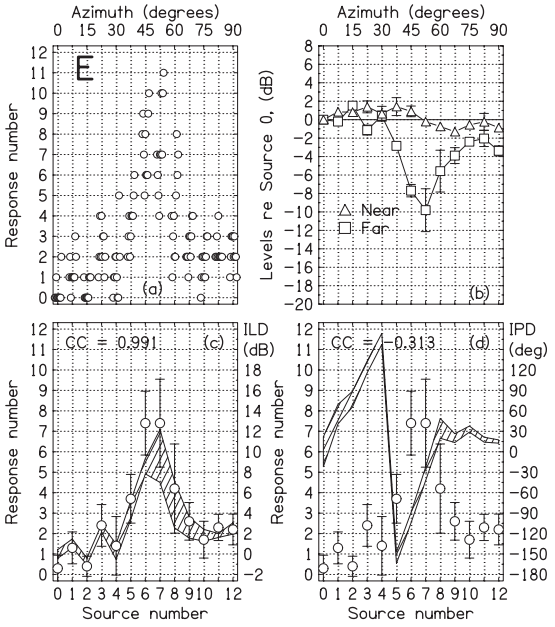


FIG. 4. Same as Fig. 3, but for listener E.

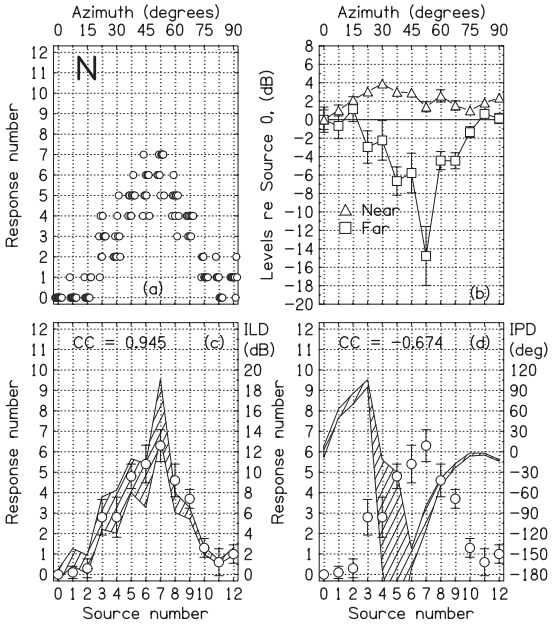


FIG. 6. Same as Fig. 3, but for listener N.

and X, respectively. The KEMAR data from Fig. 2 are the contribution to physical measurements from listener K.

In approximate agreement with the spherical-head diffraction model, the levels measured in the *far ear* showed a minimum near an azimuth of 50° (60° for listener B). The minima were of similar depth, ranging from –10 dB for listener E to –16 dB for listener B, and were all deeper than the prediction of –6 dB from the spherical-head model.

Although one might expect the levels in the near ear to be similar across the listeners because of the simple geometry, they were actually quite different. Two listeners, M and X, followed the prediction of the spherical-head model in that the levels tended to increase monotonically with azimuth, although the levels for these listeners rose by about 10 dB, to be compared with only 3 dB for the model. All the other listeners, including K, showed near-ear levels that were rather flat functions of azimuth, with some tendency to decrease near 70°.

2. ILD analysis

The ILDs shown in the (c) panels of Figs. 3–7 are the differences between near and far-ear levels. For all listeners, the ILDs have a peak where the far-ear level was a minimum. Peaks are in the range of 20–23 dB for listeners B, M, and X. They are in the range of 10–16 dB for listeners E, K, and N. All peak ILDs are greater than the value of 8 dB

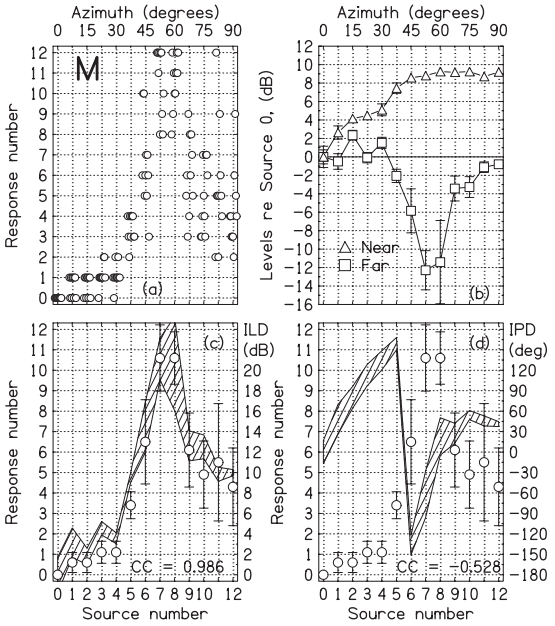


FIG. 5. Same as Fig. 3, but for listener M.

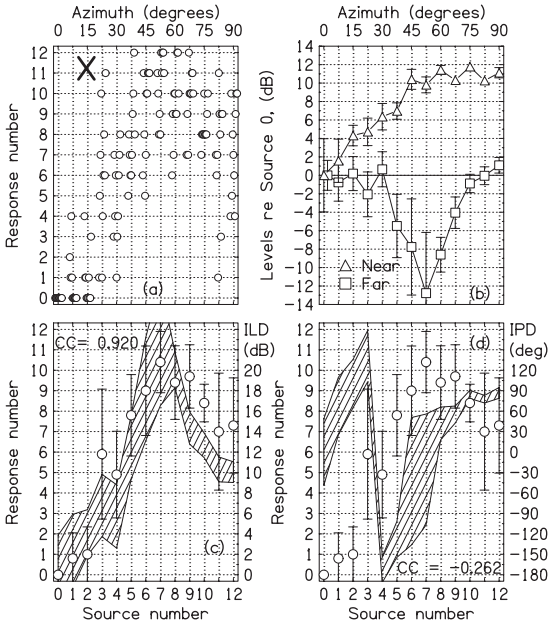


FIG. 7. Same as Fig. 3, but for listener X.

predicted by the spherical-head model. The discrepancy is likely to be a result of the torso, which is not included in the model.

The ILDs in the (c) panels are plotted as hatched polygons, 2 standard deviations in height. Given the reliability of the microphone method indicated by the KEMAR test, the standard deviations are likely caused by listener motion. Some of the deviation occurred as the listener was reseated between the first and second runs.

a. Listener responses The identification data reported here are based on the ten trials for each loudspeaker. The data are presented along with the physical measurements in Figs. 3–7 for the five listeners. In the (c) panels, the identification responses are superimposed on the ILD plots. No attempt was made to plot the listener responses in a way that would coincide with the polygons depicting measured ILDs for individual listeners. The scaling rule for drawing the plots was 7.5° of response angle for 2 dB of ILD for all the listeners. Nevertheless, the responses and ILDs overlap rather well.

The correlation (Pearson product moment) between the average response and the average ILD, computed over the 13 sources, is given in the plots. Over the five listeners, the average correlation was 0.96 (sd=0.03). These high correlations clearly suggests that localization decisions were strongly influenced by the ILD.[3]

b. Sensitivity to ILD For all listeners, the peak identification response occurred for the source (or sources) that led to the largest ILD. Listeners B, M, and X, with peak ILDs greater than 20 dB, all made peak responses greater than 10. Listeners E and N, with peak ILDs less than 16 dB, both made peak responses less than 8. The corresponding peak ILDs and peak responses suggest a common-sensitivity hypothesis. According to this hypothesis, all listeners have similar internal scales relating ILDs to azimuths, and listeners E and N gave smaller responses, simply because they experienced smaller ILDs.

To test the common-sensitivity hypothesis, the individual responses R were plotted as a two-parameter function, $R=A[1-\exp(-\mathrm{ILD}/C)]$. The initial slope of this function is A/C, and it serves as a measure of sensitivity to ILD. The largest slope was for listener B, 10°/dB. The smallest was for listener N, 5°/dB. The data for listener M did not have negative curvature and could not be fitted with the function. Listeners B, E, and X had similar slopes of 10°/dB, 7.6°/dB, and 7.1°/dB, respectively. Therefore, the common-sensitivity hypothesis could account for the small peak response made by listener E. By contrast, the small peak response for listener N reflects a relative insensitivity to stimulus ILD.

3. IPD analysis

The IPDs in the (d) panels of Figs. 3–7 are also plotted as polygons, again with overall heights showing 2 standard deviations in the measurements. In agreement with the spherical-head model, the IPDs all cross the 180° line for azimuths between 30° and 45°. There is one exception; the crossing occurs between 22° and 30° for listener X. Listener X also had the largest peak ILD and the widest head. The correlation between interaural parameters and head size was made clear in the scaling study by Middlebrooks (1999a, 1999b).

The identification responses from the (c) panels are repeated on the (d) panels for comparison with measured IPD. As shown by the correlation coefficient (cc) values on the (d) panels, the correlation between the responses and IPDs was negative for all the listeners. That negative correlation suggests that identification judgments were not much influenced by IPDs.

4. Individual-ear analysis

The strong correlation between ILD and response for all the listeners points to the ILD as a dominant cue for localization. However, there is additional localization information in the individual levels in left and right ears, and listeners might possibly benefit from that. A simple way to combine ILD with individual ear levels is, first, to observe that the nonmonotonic ILD presents the listener with an ambiguity and, second, to assume that individual ear levels might resolve that ambiguity. We called that strategy the "supplemental strategy."

a. Supplemental strategy We studied the supplemental strategy by beginning with the ILDs of all the sources to the right of the ILD peak, as shown in part (c) of Figs. 3–7. For each of these source azimuths to the right of the peak, there corresponds at least one azimuth to the left of the peak (not necessarily at a source location) for which the ILD is the same. We asked whether either the near-ear levels or the far-ear levels (usually interpolated) for those ILD-ambiguous azimuths differed by more than 1 dB. If so, we assumed that the individual ear difference could resolve the ILD ambiguity. An ideal listener with complete information could successfully employ such a strategy.

Summed over the five listeners in Experiment 1, there were 24 ILD-ambiguous sources. The supplemental strategy successfully resolved the ambiguity for 19 of these. Therefore, this hypothesis predicts that Experiment 1 should show only a few non-monotonic identification responses, contrary to observation. Apparently, human listeners are not able to use all the information assumed by the supplemental strategy.

b. Separate-ears strategy An alternative strategy, the "separate-ears strategy," involves only the levels in the two ears separately. A simple separate-ears strategy begins by assuming that the listener determines source location based on the level in the far ear because the far ear shows the stronger azimuth dependence. However, as shown in parts (b) of Figs. 3–7, the far-ear level is ambiguous because of the minimum in the level function. For every source azimuth to the right of the minimum, there is at least one azimuth (not necessarily at a source location) to the left of the minimum for which the far-ear level is the same. We assumed that a listener would try to resolve the ambiguity between two azimuths by considering the levels in the near ear. If the near-ear level for the larger azimuth is at least 1 dB greater than the near-ear level for the smaller azimuth (as it is for all but

TABLE I. Correlation between the source number and the listener's response in four identification experiments. Experiments 1 and 2 used sine waves. Experiment 2 included training and feedback. Experiment 4 used AM. Experiment 5 used narrow-band noise.

Listener	Expt.1	Expt.2	Expt.4	Expt.5
B	0.49	0.49	0.98	0.91
E	0.29	0.49	0.82	0.86
M	0.64	0.71	0.63	0.67
N	0.20	0.28	0.99	0.99
X	0.76	0.78	0.86	0.93
Av	0.48	0.55	0.86	0.87

one of the sources in the spherical-head model), then we assumed that the separate-ears analysis would resolve the ambiguity.

Calculations using the separate-ears strategy indicated that listeners E and N should not be able to resolve the ambiguity for any of the ambiguous sources. However, calculations for listeners M and X predicted successful resolution for four of the five ambiguous sources for each listener. Calculations for listener B were intermediate between those for E and N on the negative side, and M and X on the positive side. They predicted successful resolution for two of the four sources that were ambiguous for listener B. A glance at the first column of Table I shows that the predictions of the separate-ears strategy are ordered the same as performance in Experiment 1. The average correlation for listeners E and N was 0.25. The average correlation for listeners M and X was 0.7. The correlation for listener B was intermediate; a value of 0.49.

The above analysis of the separate-ears strategy, based on precise rules, does not differ from the impression that one gets merely by looking carefully at the (b) parts of Figs. 3–7. Listeners M and X exhibited a large, approximately monotonic increase in near-ear level with increasing azimuth. Other listeners, especially E and N, did not. The problem with the separate-ears strategy as a hypothesis is that it predicts mostly monotonic responses for listeners M and X, but these listeners gave non-monotonic responses like those of the other listeners.

What can be said in conclusion is that all listeners were misled by the non-monotonic ILD, but some listeners may have found it possible to use the information in separate ears to improve their decisions. Evidence for some use of separate-ear information comes from the fact that the predictions of the separate-ears strategy correlate rather well with performance in Experiment 1.

III. EXPERIMENT 2: INFORMED IDENTIFICATION

Because listeners obviously were confused by the stimuli in Experiment 1, Experiment 2 attempted to teach listeners to make better use of the available cues. There were two phases: training and feedback (T/F). The training phase was incorporated into the experiment runs. Each experiment run consisted of two data-collection passes through the 13 sources where the order of presentation of the sources was randomized on a pass. Prior to each of the two passes, there was a 1 min training pass, in which the sources were presented in ascending numerical order. The listener was alerted to the start of a training pass and knew what to expect. Each listener completed five runs, a total of ten data passes (by contrast, Experiment 1 had two runs of five passes).

A simple form of feedback was given after each trial as data were collected—one pilot lamp if the source had been in the range 0–6, the other pilot lamp if the source had been in the range 7–12. This binary feedback was adequate to resolve almost all of the ILD ambiguity.

A. Results

Table I shows the correlation between the listener's response and the true source number for the ten passes of Experiment 1 (no T/F) and the ten passes of Experiment 2 (with T/F). It shows modest improvement in Experiment 2 for three of the five listeners. To better determine whether identification accuracy benefited from T/F, the first five passes in Experiment 2 were compared with the first five passes in Experiment 1. Also, the second five passes in Experiment 2 were compared with the second five passes in Experiment 1. To determine whether continued T/F led to improved accuracy, the first and second five passes were compared within Experiment 2. The statistic used to assess accuracy was again the correlation between the listener response and the source number.

For the first five passes, the correlation was improved by T/F for three of the five listeners, and the mean increased from 0.46 to 0.52. For the second five passes the correlation improved for four of the five listeners, and the mean increased from 0.48 to 0.56. The correlation on the first pass of Experiment 2 was 0.52, and on the second pass, it was 0.56, a difference of 0.04 that might be attributed to learning.

The increases in correlations attributable to T/F, as reported above, were clearly modest at best. By contrast, correlations between responses with and without T/F were high. A round-robin comparison of first and second passes led to four correlations that ranged from 0.86 to 0.88. The conclusion of Experiment 2, incorporating training and feedback, was that listeners did learn from the experience, but they did not learn much. However, the training in Experiment 2 was not extensive, and it is possible that further training, or different training, might have been more effective.

IV. EXPERIMENT 3: DISCRIMINATION

Experiment 3 was similar in its setup to Experiments 1 and 2. The stimuli were also similar, except that the two tones on each trial came from different loudspeakers. Thus, Experiment 3 had the form of a classic two-source two-interval discrimination experiment intended to determine acuity. From the listener's perspective, the tones of a trial moved from left to right or moved from right to left along the arc. The listener's task was to report the direction of motion by means of push buttons.

A. Method

With 13 loudspeakers, there were 78 ways to choose different pairs. To obtain adequate statistics, it was necessary

to limit this number and select pairs. For each experimental run, six pairs of loudspeakers were selected based on predictions from the spherical-head calculations and from the identification data for the individual listener. Two pairs were selected because the previous data predicted that discrimination responses would be correct. Two others were selected because the data predicted incorrect responses, and two others were selected expecting uncertain responses. An experimental run included ten trials for each pair, presented in random order.

There were two experimental runs with different loudspeaker pairs selected for each. Therefore, each listener discriminated among 12 pairs, normally different for different listeners. The listeners from the identification experiments, Experiments 1 and 2, were also the listeners in Experiment 3. The timing and tone levels were the same as in Experiments 1 and 2. As in Experiments 1 and 2, ILD and IPD were measured in the discrimination experiment.

B. Predicting discrimination results

It is straightforward to use the results of the identification experiments to predict the results of discrimination experiments, assuming that the same sensory process applies to both experiments. If μ_m and μ_n are the mean identification responses for sources m and n, and if σ_m and σ_n are the corresponding standard deviations, then the expected value of d' in an experiment that requires the listener to discriminate between sources m and n is

$$d'_{m,n} = \frac{\mu_m - \mu_n}{\sqrt{\sigma_m^2 + \sigma_n^2}}. \quad (1)$$

This value of d' can be converted to percent correct on a two-interval forced-choice task. Such predicted values of percent correct appear on the horizontal axis in Fig. 8(a).

C. Results

The comparison between the discrimination responses of Experiment 3 and the predictions are shown in Fig. 8(a) for the five listeners. The predictions were successful to some degree because almost all of the data points fell into the upper-right and lower-left quadrants of the plot. Points in the low-left quadrant correspond to systematic discrimination errors and show that the confusions in identification caused by the bright spot reappear in discrimination.

The shape of the predictions in Fig. 8(a) is different from the responses. The predictions were on a continuum from 0% to 100% correct. The responses tended to be bimodal. Evidently listeners were much more certain about whether one source was to the right or left of another than they were about the absolute locations of the two sources involved. Informally, listeners remarked that the discrimination task seemed easier than the identification task.

The discrepancy between the actual and predicted results in Fig. 8(a) is unlikely to be the result of technical error. As per Moore *et al.* (2008), who obtained a good correspondence between identification and discrimination, values of μ and σ in Eq. (1) were taken from source-dependent behavioral data. The problem is likely that the values of σ were

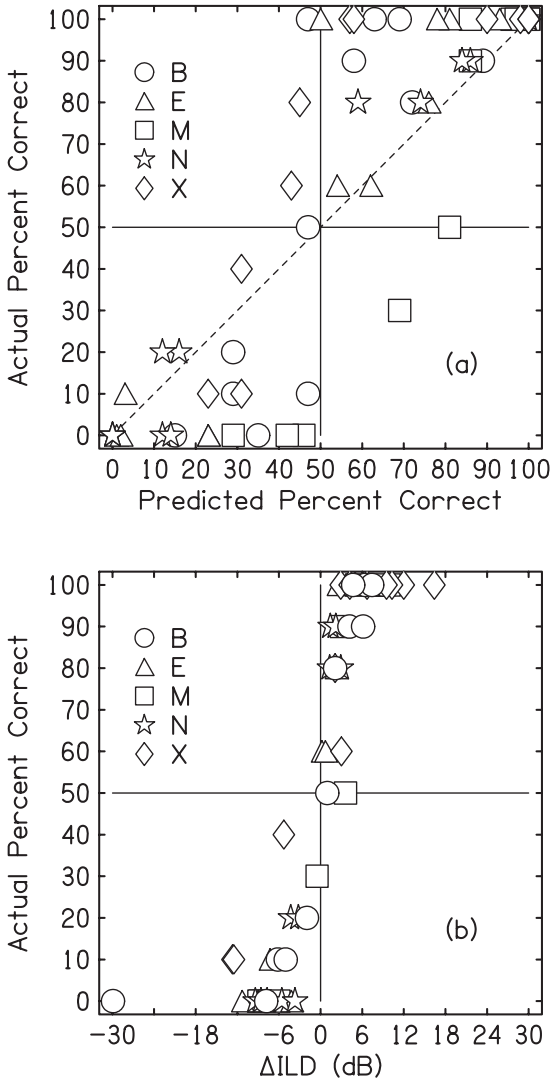


FIG. 8. Experiment 3, Discrimination. (a) The predicted percent correct and the actual percent correct are plotted against each other for the five listeners. Predicted values were based on individual identification data from Experiment 1. (b) Same as part (a) except that the actual percent correct is plotted against the change in ILD between the greater and lesser source numbers.

obtained from identification experiments that covered a wide angular span. As noted by Shelton and Searle (1978), the standard deviation in identification experiments grows with both azimuth and with angular span. The effect was modeled by Hartmann and Rakerd (1989) as a sensory bias, which may rove in an experiment that is extensive over both azimuth and time. The increased standard deviation in identification would lead to a continuous form of the predicted results, as in Fig. 8(a).

In addition to sensory bias, the identification experiment may have suffered from a peculiar form of response bias. Subjects who know that there are 13 sources may expect to perceive all 13 over the course of an experiment. When many trials have occurred, and all of them point to only half of the sources, subjects may try to compensate. Response bias like this is one way to account for the scatter of data in the (a) parts of Figs. 3–7, particularly for listeners B, M, and X. These are the listeners whose data depart most dramatically from the 45° line in Fig. 8(a).

An alternative to a discrimination prediction based on identification data is a prediction based on physical data. Figure 8(b) shows the comparison between the discrimination responses and the change in ILD between the two sources. A comparison with Fig. 8(a) indicates that the change in ILD, especially the sign of the change, is a much better predictor of discrimination than is the identification performance. Only when the change in ILD was less than about 4 dB did the listeners tend to give ambiguous responses.

V. EXPERIMENT 4: AMPLITUDE MODULATION

Experiment 4 was an identification experiment like Experiment 1, except that the tone was given 100% amplitude modulation (AM) with a modulating frequency of 100 Hz. The temporal window for the AM tone was the same as for Experiment 1.

The motivation for the experiment was that the envelope of AM tones provides a temporal pattern that enables the binaural system to use ITDs, even at a relatively high frequency such as 1500 Hz (Henning, 1974; McFadden and Pasanen, 1976; Stellmack *et al.*, 2005). The 100-Hz modulation frequency was well below the 800-Hz limit found in the superior olive by Joris and Yin (1998) and in references contained therein. The 200-Hz bandwidth is close to the auditory filter width at 1500 Hz, measured by Glasberg and Moore (1990). The listeners for previous experiments were listeners in Experiment 4.

A. Results

The results of Experiment 4 are shown for the five listeners by individual plots, showing response number vs source number in Fig. 9(a). The figure shows that the responses for listeners B, N, and X became monotonic functions of azimuth or nearly so. However, the responses for listeners E and M remained non-monotonic. As shown in Table I, the correlation between response and source numbers was 0.98 or greater for listeners B and N. Although listener E's responses remained non-monotonic, the correlation increased considerably, compared to Experiment 1, an increase from 0.29 to 0.82. By contrast, the addition of AM led to no change in the responses of listener M by any of our measures.

VI. EXPERIMENT 5: NARROW-BAND NOISE

In Experiment 5, the sine tone of Experiment 1 was replaced by a narrow-band noise (NBN). Again, the motivation for the experiment was to determine if temporal features in the envelope of the stimulus could provide an ITD cue that would enable localization at large azimuth. It might be expected that NBN would be more effective than AM because of the wider range of envelope-variation frequencies.

The experiment used equal-amplitude, random-phase noise with a spectrum consisting of 201 components spanning the range 1400–1600 Hz. Therefore, the bandwidth was the same as for Experiment 4. There were actually 13 different noises, based on different phase randomizations, selected

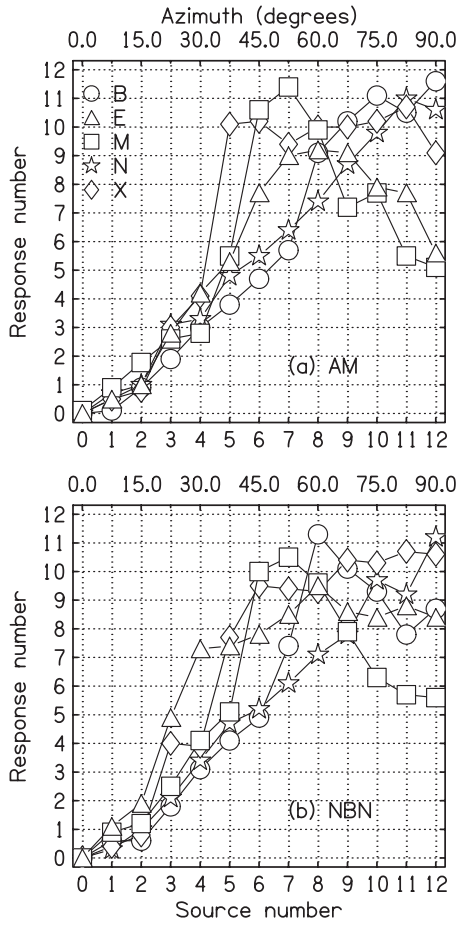


FIG. 9. (a) Experiment 4, AM: Listener identification responses on the scale 0–12 to sources as numbered on the horizontal axis for five listeners. The stimulus was a 1500-Hz sine tone amplitude modulated by a 100-Hz sine tone. (b) Experiment 5, NBN: Same as (a) except that the stimulus was a 200-Hz band of noise centered on 1500 Hz.

randomly for each experimental trial. The temporal window for the noise was the same as for Experiment 1. The same listeners participated.

A. Results

The results of Experiment 5 (NBN) are presented in Fig. 9(b), which can be compared with Experiment 4 (AM) in Fig. 9(a). The data for listeners N and X look very similar for AM and NBN, which was the expected result. Compared to the results for AM, the results for listener E became more monotonic, and the results for listener B became less. Listener M was no more aided by NBN than by AM. His results were unchanged by any measure. Table I shows that the correlations between listener responses, and source numbers were similar in Experiments 4 and 5.

The experiments incorporating AM or NBN show that for most of the listeners, the temporal structure in the envelope can be used to solve the localization problem given an ambiguous ILD. [Experiments by Eberle *et al.* (2000) suggest that there would be little value in attempting to combine NBN with AM to obtain additional time structure.] However, for some listeners, listener M in particular, the timing information in the envelope proved to be insignificant, compared to the ILD cue.

VII. DISCUSSION AND CONCLUSION

The experiments of this article demonstrated the importance of the information, and misinformation, contained in the ILD for medium-frequency pure tones such as 1500-Hz pure tones. Experiment 1 showed that the non-monotonic ILD, predicted by the bright-spot theory and the spherical-head model, occurred for real listeners, and that it led to a parallel non-monotonic perception of azimuth in a source-identification experiment. Experiment 2 showed that the misleading effects of the non-monotonic ILD were not alleviated by training the listeners nor by giving them feedback after every trial. Experiment 3 showed that the confusion indicated in the identification experiments translated to predictable confusion in the discrimination experiments. However, the ILD values themselves were better predictors of discrimination performance than were the results of identification experiments, because the discrimination experiments showed smaller within-subject variability than the identification experiments. The increased variability in identification could be attributed to both sensory bias and response bias.

A foreshadowing of the bright-spot effects seen in Experiments 1–3 on pure tones appeared in the discrimination experiments by Mills (1958, 1972). Mills measured the minimum audible angle (MAA) as a function of reference angle and frequency with the following results.

- For reference angles in the forward direction (azimuth 0), the MAA demonstrated a familiar behavior: For low frequencies, the MAA was matched by the just noticeable difference (JND) in ITD, as measured by headphone experiments. When the frequency increased to 1500 Hz, the MAA (about 3°) was matched by the JND in ILD, indicating that the ITD had lost its effectiveness.
- When the reference azimuth increased to 30°, the MAA between 1500 and 4000 Hz increased to about 6°, but was otherwise well behaved, indicating that the ILD provided a useful cue.
- However, when the reference angle increased to 60° or 75°, the MAA went off the chart as the frequency rose to 1500 Hz. In Mills' own words, "for tones between 1500 and 2000 Hz, from sources at azimuths of more than 45°, the minimum audible angle is indeterminately large."

The anomalous behavior reported by Mills finds a ready explanation in the calculations, measurements, and perceptual consequences of the bright spot, as reported in the present article. It is exactly these frequencies and these reference azimuths, where the ILD is a decreasing function of azimuth. The more the source moves to the right, the more the listener hears it to the left. Discrimination measurements like the MAA are bound to fail under such conditions, as shown by the results in Experiment 3 above.

The results of Experiments 1–3 have implications for the duplex theory of pure-tone localization. According to the version of duplex theory by Stevens and Newman (1936), there are three frequency regions. In a low-frequency region, below about 1 kHz, localization is cued by the ITD. In a high-frequency region, above 4 kHz, localization is cued by the ILD. In a middle region, extending from about 1 to 4 kHz, neither ITD nor ILD is an effective cue. Stevens and Newman based this viewpoint on localization experiments that showed a broad peak in localization error, centered on 3 kHz. Errors were notably smaller when the frequency increased to 4 or 5 kHz. These experiments measured localization errors over the full 180° from front to back, around the right side of the listener. Therefore, these experiments encountered the angles and the frequencies where the experiments of the present article find that ILD is a non-monotonic function of azimuth. This non-monotonic behavior above 1 kHz leads to large localization errors. We suggest that Stevens and Newman would have found very different results if they had limited their localization task to azimuths within 40° of the midline. ILDs in the region from 1 to 4 kHz can be large enough to support good localization near the forward direction, with errors much smaller than reported by Stevens and Newman.

It is possible that a localization experiment that avoided the azimuthal region of non-monotonic ILD would find that the midfrequency region identified by Stevens and Newman (also observed by Mills) would disappear. Alternatively, a region of relatively poor localization might emerge centered near 1 kHz because of the peak in ILD difference limen observed by Grantham (1984). Grantham speculated that this peak reflected a greater localization utility for interaural cues at frequencies, both above and below 1 kHz. Frequencies above 1 kHz are useful because of the increased ILD; frequencies below 1 kHz are useful because latency in peripheral neurons converts an ILD into an ITD.

A different explanation for the 1-kHz peak observed by Grantham is also based on localization utility. Figure 1 suggests that near 1 kHz, non-monotonic ILDs are present at rather small azimuths, 40° to 50°. Even smaller azimuths would be expected from more realistic models of the human head. At frequencies well above 1 kHz, the non-monotonic behavior is much more dramatic, but it occurs only for larger azimuths. One can speculate that azimuths within 45° of the midline are particularly important, and that binaural development has been influenced by the unreliability of ILDs at these relatively small azimuths and at frequencies near 1 kHz.

An anonymous reviewer of this article pointed out that the non-monotonic ILD caused by the bright spot has implications for front-back localization. Normally, a listener is able to resolve front-back ambiguities by rotating the head slightly because rotation causes the ILD to change in a way that favors the ear that is approaching the source (Wallach, 1939, 1940; Perrett and Noble, 1997). However, if the ILD is a decreasing function of the azimuth, the sign of the ILD change will be reversed, thus cueing the front direction when the source is in the back and vice versa. For a frequency of 1500 Hz, this kind of reversal would be expected for all sources having a lateral angle of 55° or more (e.g., 65° azimuth vs 115°).

Experiments 4 and 5 modified the stimulus by adding temporal structure to the envelope of the tone, which significantly reduced the confusion in source identification on the

average. However, the effectiveness of the temporal structure varied. For different listeners, it ranged from completely effective to not-at-all effective.

The modulated pure tone is a stimulus with two localization cues, the ILD in the entire tone, and the ITD in the envelope modulation. Experiments 4 and 5 indicated that different listeners weight these two cues differently. This observation is consistent with a general model of sound localization, wherein different, possibly conflicting, localization cues arrive at a central processor in the nervous system where they are combined to form a localized image. The process of weighting and combining the cues is normally entirely subconscious. A model of localization or lateralization that depends on idiosyncratic central weighting of cues has long been a possible interpretation of ITD-ILD trading experiments (e.g., Hafter and Carrier, 1972). Survey experiments by McFadden *et al.* (1973) found an anomalously large sensitivity for one of these cues or the other in 50% of the population, and of these, 75% were more sensitive to the ILD. The failed attempts to retrain listeners described by Jeffress and McFadden (1971) indicate that, although the weighting is not necessarily hard wired, it is very resistant to change. The central weighting hypothesis gained practical value in connection with experiments on the localization of sound in rooms, where standing waves lead to conflicting cues (Hartmann, 1983; Rakerd and Hartmann, 1985). Such a central model of cue combination would seem to apply to the highly individualized weighting of cues observed in Experiments 4 and 5.

ACKNOWLEDGMENTS

Mr. Zachary Ryan provided important technical help in the early phases of this work, supported by the NSF REU program in the Department of Physics and Astronomy at Michigan State University. This work was mainly supported by the NIDCD of the NIH under Grant No. DC 00181.

[1]The calculations are from the low-frequency limit of the diffraction formula, where the ITD is given by $(3r/c)\sin\theta$, where r is the radius of the head, c is the speed of sound, and θ is the azimuth with respect to the forward direction (Kuhn, 1977).

[2]The classic optical bright spot is at the center of a shadow cast by the sphere on a screen. A bright spot on the dark surface of the sphere itself is caused by the same physics—constructive addition of diffracted waves at a symmetry point.

[3]Because a sine tone, as used in these experiments, conveys so little localization information, a listener's association of an ILD cue with a location in the horizontal plane occurs partly because the listener can see sources in that plane. Because of the cone of confusion, or the equivalent for real heads, many other source locations would also correspond to a given ILD cue.

Burkhard, M. D. and Sachs, R. M. (**1975**). "Anthropomorphic manikin for acoustics research," J. Acoust. Soc. Am. **58**, 214–222.

Eberle, G., McAnally, K. I., Martin, R. L., and Flanagan, P. (**2000**). "Localization of amplitude modulated high-frequency noise," J. Acoust. Soc. Am. **107**, 3568–3571.

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

Grantham, D. W. (**1984**). "Interaural intensity discrimination: Insensitivity at 1000 Hz," J. Acoust. Soc. Am. **75**, 1191–1194.

Hafter, E. R., and Carrier, S. C. (**1972**). "Binaural interactions in low-frequency stimuli: The inability to trade time and intensity completely," J. Acoust. Soc. Am. **51**, 1852–1862.

Hartmann, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle—A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.

Hartmann, W. M., and Wittenberg, A. T. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Hecht, E. (**2002**). *Optics* (Addison-Wesley, San Francisco, CA), p. 494.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Jeffress, L. A., and McFadden, D. (**1971**). "Differences of interaural phase and level in detection and lateralization," J. Acoust. Soc. Am. **49**, 1169–1179.

Joris, P. X., and Yin, T. C. T. (**1998**). "Envelope coding in the lateral superior olive III. Comparison with afferent pathways," J. Neurophysiol. **79**, 253–269.

Kelly, W. R., Shirley, E. L., Migdall, A. L., Polyakov, S. V., and Hendrix, K. (**2009**). "First and second order Poisson spots," Am. J. Phys. **77**, 713–720.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.

Kuhn, G. F. (**1987**). *Directional Hearing*, edited by W.A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.

McFadden, D. (**1981**). "The problem of different interaural time differences at different frequencies," J. Acoust. Soc. Am. **69**, 1836–1837.

McFadden, D., Jeffress, L. A., and Russell, W. E. (**1973**). "Individual differences in sensitivity to interaural differences in time and level," Percept. Mot. Skills **37**, 755–761.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Am. Stat. Assoc. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Mills, A. W. (**1972**). *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York).

Moore, J. M., Tollin, D. J., and Yin, T. C. T. (**2008**). "Can measures of sound localization acuity be related to the precision of absolute location estimates?," Hear. Res. **238**, 94–109.

Perrett, S., and Noble, W. (**1997**). "The effect of head rotations on vertical plane sound localization," J. Acoust. Soc. Am. **102**, 2325–2332.

Rakerd, B., and Hartmann, W. M. (**1985**). "Localization of sound in rooms II: The effects of a single reflecting surface," J. Acoust. Soc. Am. **78**, 524–533.

Rschevkin, S. N. (**1963**). *A Course of Lectures on the Theory of Sound*, translated by P. E. Doak (Pergamon, New York/McMillan, New York).

Shelton, B. R., and Searle, C. L. (**1978**). "Two determinants of localization acuity in the horizontal plane," J. Acoust. Soc. Am. **64**, 689–691.

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Discrimination of interaural phase differences in the envelopes of sinusoidally amplitude modulated 4-kHz tones as a function of modulation depth," J. Acoust. Soc. Am. **118**, 346–352.

Stevens, S. S., and Newman, E. B. (**1936**). "The location of actual sources of sound," Am. J. Psychol. **48**, 297–306.

Wallach, H. (**1939**). "On sound localization," J. Am. Stat. Assoc. **10**, 270–274.

Wallach, H. (**1940**). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Speech localization in a multitalker mixture[a)]

Norbert Kopčo[b)]
Department Of Cybernetics and AI, Technical University of Košice, Košice 04001, Slovakia

Virginia Best and Simon Carlile
School of Medical Sciences, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 27 June 2009; revised 24 November 2009; accepted 15 December 2009)

An experiment was performed that measured, for the frontal audio-visual horizon, how accurately listeners could localize a female-voice target amidst four spatially distributed male-voice maskers. To examine whether listeners can make use of *a priori* knowledge about the configuration of the sources, performance was examined in two conditions: either the masker locations were fixed (in one of five known patterns) or the locations varied from trial to trial. The presence of maskers disrupted speech localization, even after accounting for reduced target detectability. Averaged across all target locations, the rms error in responses decreased by 20% when *a priori* knowledge about masker locations was available. The effect was even stronger for the target locations that did not coincide with the maskers (error reduction of 36%), while no change in errors was observed for targets coinciding with maskers. The benefits were reduced when the target-to-masker intensity ratio was increased or when the maskers were in a pattern that made it difficult to make use of the *a priori* information. The results confirm that localization in speech mixtures is modified by the listener's expectations about the spatial arrangement of the sources.
 [DOI: 10.1121/1.3290996]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.71.Rt [JCM] Pages: 1450–1457

I. INTRODUCTION

It is known that spatial factors play a role in the ability of listeners to understand speech in noisy or complex listening environments. For example, acoustical advantages are provided by the spatial separation of speech from competing noise (Zurek, 1993; Bronkhorst, 2000). In the case of multiple competing talkers, it is believed that spatial differences also enable the correct "sorting" of the acoustic mixture into different sources and enable listeners to direct attention selectively to one source to enhance its processing (Yost *et al.*, 1996; Freyman *et al.*, 1999; Brungart, 2001; Shinn-Cunningham, 2008). In addition, the ability to rapidly locate a talker of interest in order to focus on them visually is clearly an important aspect of human communication. Despite the various roles of spatial hearing in dealing with competing speech sources, we know surprisingly little about how accurately listeners can localize in speech mixtures.

Under ideal conditions, humans can localize single broadband sounds within a few degrees of accuracy (Mills, 1958; Wightman and Kistler, 1989; Carlile *et al.*, 1997; Best *et al.*, 2005). When presented against a background of noise, localization does not suffer until the signal-to-noise ratio is negative, due in part to the fact that the *detection* of the target becomes compromised (Good and Gilkey, 1996; Good *et al.*, 1997; Abouchacra *et al.*, 1998; Lorenzi *et al.*, 1999). Few studies have examined more complex situations involving multiple talkers (for a summary see Faller and Merimaa, 2004). Hawley *et al.* (1999) measured localization of a known target sentence in the presence of one to three unknown sentences. Their task was a 1-of-7 loudspeaker identification (30° spacing), and they found that performance was relatively accurate (around 70%) and not affected significantly by the number or configuration of the maskers. In a similar paradigm with one to four distractors, Drullman and Bronkhorst (2000) found poorer performance (around 50%) for a 1-of-5 loudspeaker identification task (45° spacing). Finally, Simpson *et al.* (2006) required listeners to detect and localize a known word in a mixture of one to five synchronous words. The number and configuration of the maskers were varied from trial to trial. The authors found that localization errors increased systematically with the number of maskers even when miss trials (where the subject could not detect the target) were excluded from the analysis.

Few previous studies asked whether providing *a priori* information that can be used to direct automatic or strategic attention can improve sound localization (Spence and Driver, 1994; Sach *et al.*, 2000; Kopčo *et al.*, 2001). In these studies, the effect of cuing the target location was small: improvements in reaction times were observed (Spence and Driver, 1994), but little (Sach *et al.*, 2000) or no (Kopčo *et al.*, 2001) improvement in localization accuracy was found. A possible explanation for these weak effects is that the target was presented in isolation, and thus a lot of redundant information about its location was available. More complex scenes, in which sources compete for attention, may be more likely to reveal effects of *a priori* information.

The current study aimed to examine, in a realistic, complex listening situation containing competing sound sources, whether *a priori* knowledge of the location of maskers might

[a)]Portions of this work were presented at the 157th meeting of the Acoustical Society of America.
[b)]Author to whom correspondence should be addressed. Electronic mail: kopco@bu.edu

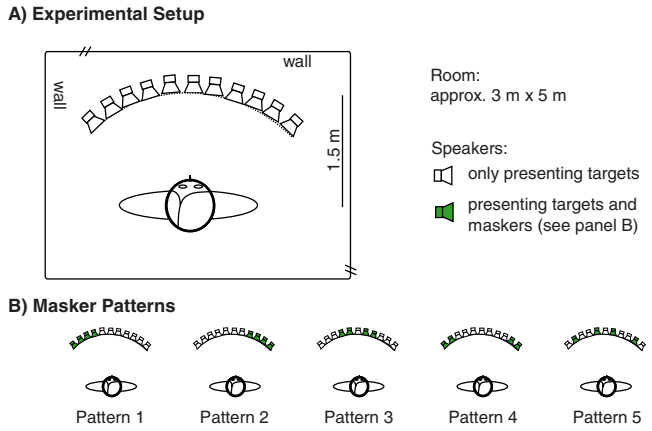


FIG. 1. (Color online) (A) Eleven loudspeakers evenly spaced in front of the listener (10° separation) were used to present stimuli. (B) Maskers were presented from loudspeaker locations arranged into one of five masker patterns. Each pattern had four maskers, presented concurrently with the target.

modify how they affect target localization performance. To this end, we measured the accuracy with which listeners can localize a speech target presented from a random location in the presence of four speech maskers. While all maskers were spatially separated from each other, targets could occur anywhere (including at masker locations). The maskers were presented in one of five masker patterns (see Fig. 1). In separate blocks, we either varied the masker pattern randomly from trial to trial (Mixed condition) or kept one of the masker patterns fixed throughout a block, providing the subject with information about the maskers' locations at the beginning of the block (Fixed condition). We expected that subjects would be able to use the masker location information in the Fixed condition to improve their localization performance (with regard to Mixed condition). For example, by actively suppressing the masker locations or actively attending away from masker locations, performance might improve for targets at nonmasker locations (but perhaps worsen for targets colocated with maskers).

The second goal of this study was to examine whether the extent to which *a priori* information can be utilized is influenced by the complexity of the masker distribution. Therefore, we included two simple masker configurations (patterns 1 and 2 in Fig. 1, in which all masker locations and all nonmasker locations were clustered together), two intermediate configurations (pattern 3, in which the masker locations were in one cluster while the nonmasker locations were in two clusters, and pattern 4, with the opposite arrangement), and one complex configuration (pattern 5, in which both the masker and the nonmasker locations were approximately evenly distributed). We expected the benefit of *a priori* knowledge to be larger for the simpler configurations, based on the assumption that spatial attention or masker suppression can be more efficiently applied to a single region than to multiple regions.

II. METHODS

A. Subjects

Seven subjects participated, one female and six males between the ages of 18 and 50 years. All had normal hearing by self-report and gave informed consent as required by the University of Sydney's Human Research Ethics Committee.

B. Environment, stimuli, and setup

The experiment took place in an empty office of dimensions $3 \times 5 \times 2.5$ m (width × length × height). The room had carpet on the floor, a concrete ceiling, and plasterboard on the lateral and back walls. The front wall was exposed brick and contained a large window that was filled with thick wool matting to reduce reflections and block out most of the incoming light; as a result visibility of the loudspeakers was minimal. Eleven loudspeakers on stands were positioned with a spacing of 10° to form a horizontal arc of radius 1.5 m [Fig. 1(A)] at the level of an average listener's ears when standing (1.6 m). The target was presented from 1 of the 11 loudspeakers at random. On masker trials, four simultaneous maskers were arranged in one of five configurations [Fig. 1(B)]. Listeners were aware that targets could fall on masker locations.

Speech materials were taken from a corpus of monosyllabic words recorded at Boston University's Hearing Research Center (Kidd *et al.*, 2008a). The target was the word "two" spoken by one of the female voices in the corpus. Maskers were nondigit words spoken by the eight male talkers in the corpus, and included names (e.g., "Jane"), verbs (e.g., "found"), adjectives (e.g., "red"), and nouns (e.g., "toys"). The four masker words were drawn randomly from this set with the constraints that they were four different words spoken by four different male voices. On catch trials (see below), the target was replaced by another randomly chosen masker word.

The maskers were all longer in duration than the target, and because they were spoken by male talkers they generally had a broader spectrum than the target. Thus, they did not provide substantial spectral or temporal gaps in which the listener could have a good "glimpse" at the target (as would be the case if the target and maskers were sentences containing natural silent breaks). On the other hand, keeping the target word fixed made the task of identifying the target easier than in natural situations in which the word spoken by the target speaker varies continuously (as do acoustic, phonetic, prosodic, and other characteristics of the utterance). This as well as the difference in gender between the target and the maskers enabled listeners to focus on localizing the target accurately without any ambiguity about which sound to localize.

Target words were presented at a level of approximately 60 dB SPL(A). Maskers were all equal in level, but presented at one of two levels relative to the target in order to vary the difficulty of the task. In the easier task, each masker was equal in level to the target [target-to-masker ratio (TMR) of 0 dB]; in the more difficult task, each masker was 5 dB louder (TMR of −5 dB).

The experiment was run in MATLAB on a PC-compatible control computer. On each trial, the appropriate stimuli were loaded from files stored on the computer hard disk (at a sampling rate of 48 kHz) and sent via a multichannel soundcard (RME Fireface 400), D/A converter (Apogee DA-16x),

and amplifier (Ashley Powerflex 6250), to Tannoy V6 loudspeakers.

Subjects indicated their responses by pointing their head in the perceived direction of the target and pressing a handheld response button. A headtracker (Intersense IC3) mounted on a plastic headband was used to measure the orientation of the head at the time of response.

C. Procedures

Before a session, the experimenter positioned the subject such that he/she was in the center of the loudspeaker array with his/her head pointing to 0° azimuth, and this location was recorded by the headtracker as the reference position. Before the stimulus was played on each trial, the subject was required to orient their head to this position and feedback was given by way of a small light-emitting diode display positioned above and behind the speaker array.

In runs containing maskers, it was expected that there would be a number of trials in which the listener would not be able to detect the target, and localization responses would either be to one of the masker locations or to some random location. To avoid these trials affecting the localization data, listeners were instructed to give a specific response if they did not detect the female target (miss trials). This response was to point to a location directly above the head—a response that was easily distinguished from regular localization responses that all had an elevation component on or near 0° (on the audiovisual horizon). To ensure that listeners were following this instruction, a number of catch trials were included in which the target was replaced by another random male masker, and thus false alarm rates could be monitored.

Control runs consisted of 55 trials (5 trials per target location). In masker runs, five catch trials were also included and thus these runs were 60 trials long. Each session consisted of 12 runs. The first and last of these were control runs with no maskers present. In five runs the masker pattern was kept fixed (Fixed) for the duration of the run. Each run used one of the five masker patterns, and the pattern was indicated at the start of the run by presenting a recording of the phrase "fixed maskers" sequentially at each of the four masker locations. In the remaining five runs the masker pattern was randomly chosen on each trial (Mixed) and the run was preceded by a presentation of the phrase "mixed maskers." The Fixed and Mixed runs were interleaved. Each subject completed four sessions, two at each TMR.

III. RESULTS

A. Control data

Figure 2 plots rms errors relative to the mean response to a single unmasked target as a function of location. Separate lines are shown for the control runs of the 0 dB TMR sessions and of the −5 dB TMR sessions (although the target level did not change). rms errors were consistent across the two masker sessions, growing with target laterality from about 2° to 5°. Also, there was a slight asymmetry in errors; on the left-hand side, rms errors grew approximately linearly with target eccentricity, whereas on the right-hand side the growth was initially steep (resulting in noticeably larger errors for the +30° and +40° targets compared to the −30° and −40° targets) and then leveled off (at +50°). This asymmetry could be related to minor asymmetries in the experimental setup or the room acoustics (even though the room was mostly left-right symmetric) or it could suggest that there is some perceptual asymmetry in speech localization.

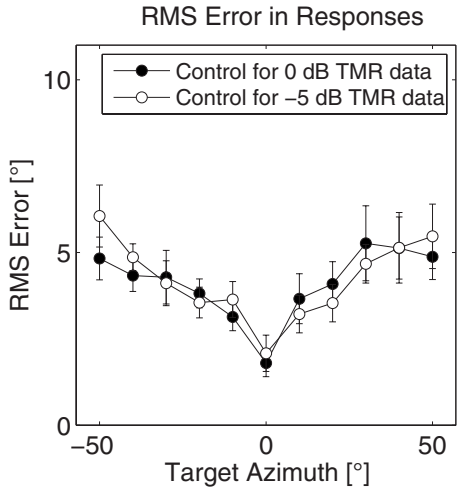


FIG. 2. Localization performance in the control condition with no maskers. Plotted are across-subject averages (±1 SEM) of the rms error as a function of the target location. Results are plotted separately for the two TMR sessions even though the target level was identical.

B. Miss rates and false alarms

Table I shows mean miss rates and false alarm rates for the two masker conditions (Fixed and Mixed). Miss rates and false alarm rates were larger at −5 dB TMR than at 0 dB TMR but were relatively low overall. It was especially important that the false alarm rate was low, as this was our indicator that subjects were reliable at indicating that they did not hear the target (and thus that they were unlikely to give random localization responses). Note that only five catch trials were included in each experimental run, and thus the false alarm rate of 20% observed at the −5 dB TMR corresponds to just one false alarm per run. It is hard to estimate the impact this nonzero false alarm rate has on performance. However, importantly, miss rates and false alarm rates were similar for Fixed and Mixed conditions, meaning that differences between these conditions were unlikely to be attributable to differences in detection criteria.

TABLE I. Detection performance averaged across subjects, masker patterns and masker locations. Miss rate shows the percentage of trials on which the target was presented but not heard (out of 55 trials per run). False alarm rate shows the percentage of catch trials (out of five trials per run) on which no target was presented but the subject gave a localization response indicating that he/she heard a target.

	Miss rate (%)		False alarm rate (%)	
Masker condition	TMR 0 dB	TMR −5 dB	TMR 0 dB	TMR −5 dB
Fixed	1	8	5	21
Mixed	1	9	8	19

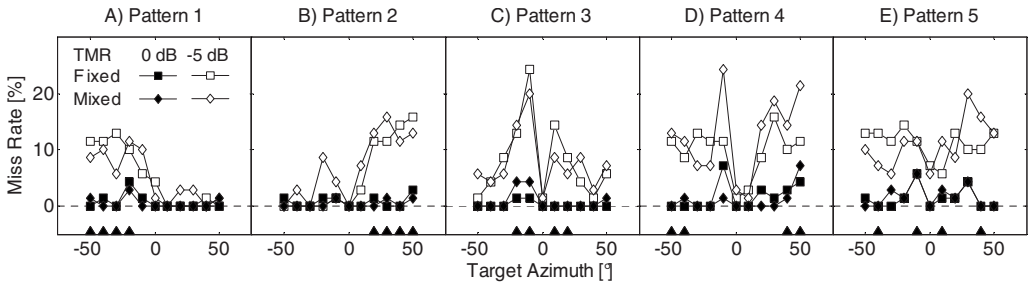


FIG. 3. Across-subject average of the miss rate (percentage of trials on which target was presented but not heard) as a function of target location. Each panel shows data for one masker pattern (masker locations indicated by the filled triangles along the abscissa), separately for all combinations of the Fixed and Mixed conditions and of the 0 and −5 dB TMR. Mean miss rates collapsed across target locations are shown in Table I.

Figure 3 shows the miss rate as a function of target location for TMRs of 0 dB (filled symbols) and −5 dB (open symbols). The five panels show data for the five different masker patterns. Misses were more common for targets falling on masker locations (or within one loudspeaker from a masker) and very rare elsewhere. This effect, and its exaggeration at the lower TMR, is likely to be an effect of energetic masking, where colocated maskers simply reduce the audibility of the target.

C. Effects of masking and *a priori* information on rms errors

Figure 4 shows the effect of maskers on rms errors for each target location. For each subject, rms errors in the control condition (see Fig. 2) were subtracted from rms errors in the different masker conditions[1] and plotted are the across-subject means of these differences. The two rows show data for the two TMRs, and the five columns show data for the five different masker patterns.

The effect of masking on rms errors depended in a complex way on all four parameters manipulated in this study, resulting in increases as large as 15° [patterns 4 and 5, Figs. 4(I) and 4(J)]. Overall, the presence of maskers always resulted in an increase in error, even at 0 dB TMR (all data points are positive in Fig. 4). It appears that the rms errors tended to increase most at target locations that corresponded to masker locations. Moreover, the largest increases occurred for masker patterns 4 and 5, the patterns in which the maskers were distributed so that they did not form one group. Lowering the TMR resulted in an approximately constant increase in rms error across all patterns.

Increases in rms error were not perfectly left-right symmetric for the symmetrical patterns 3–5 or for patterns 1 vs 2 which are mirrored versions of one another. The asymmetry appears to parallel that seen in the control data (Fig. 2), where errors are slightly higher on average for targets on the right. It is difficult to determine from these results whether such asymmetries are perceptual or reflect asymmetries in the setup and the room. The main effects of interest here (like the differences between the Fixed and Mixed performance) did not appear to be influenced by this asymmetry.

Figure 4 shows that the effect of maskers on rms errors also depended on whether the masker locations were fixed or mixed within a run. For example, for pattern 1 at the poorer TMR [Fig. 4(F)], rms errors were larger in the Fixed condition when the target was presented from the left, but were larger in the Mixed condition when the target was presented from the right. In general there was a tendency for a transition such as this to occur at or near the boundary between masker regions and nonmasker regions.

Figure 5(A) provides a summary by showing the increase in the rms error in localization responses (relative to

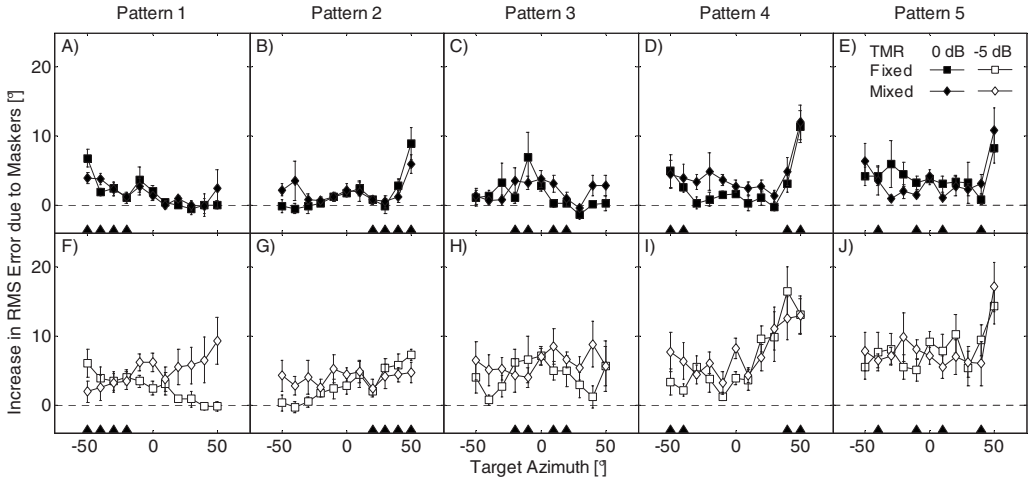


FIG. 4. Across-subject average (±1 SEM) of the increases in response rms errors (*re.* the control condition) as a function of the target location. Each column of panels shows the Fixed and Mixed condition data for one masker pattern (masker locations indicated by the filled triangles along the abscissa) and for the TMR of 0 dB (upper panels) and −5 dB (lower panels).

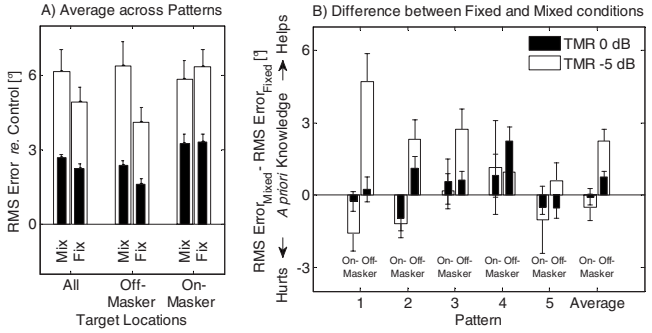


FIG. 5. (A) Effect of the maskers on localization accuracy shown as the increase in rms error in the responses to masked targets (*re.* rms error in the no-masker control condition) averaged across all patterns and across either all target locations (all), across the target locations from which no masker was presented for a given masker pattern (off-masker), or across the target locations from which a masker was presented for a given pattern (on-masker). Data are plotted separately for the Fixed and Mixed conditions and the two TMRs. (B) Effect of *a priori* knowledge on localization accuracy. The difference between rms errors in the Fixed and Mixed conditions is plotted separately for the on-and off-masker locations of each pattern and for the across-pattern average. All bars show the across-subject average (±1 SEM).

the rms error in the control condition) averaged across masker patterns. Data are averaged across all target locations (All), across the locations at which the target was not colocated with a masker (off-masker), and across the locations at which the target was presented with a colocated masker (on-masker). Separate bars represent the Fixed and Mixed conditions and the two different TMRs.

Averaged across all target locations, the reductions in the rms error in the Fixed condition with regard to the Mixed condition were 15% at 0 dB TMR and 20% at −5 dB TMR [see the filled and open "All" bars in Fig. 5(A)]. When only the off-masker target locations were considered, the effect of *a priori* knowledge was even larger, reducing the rms errors by approximately 31% at 0 dB TMR and by approximately 35% at −5 dB TMR ["Off-Masker" bars in Fig. 5(A)]. On the other hand, the availability of *a priori* information had a modest effect on the on-masker targets, increasing the rms errors by approximately 2% at 0 dB TMR and by approximately 9% at −5 dB TMR ["On-Masker" bars in Fig. 5(A)].

Figure 5(B) evaluates the effect of *a priori* knowledge directly by showing the difference between the Mixed and Fixed condition rms errors as a function of the masker pattern (including the across-pattern average), separately for the on- and off-masker locations. At the off-masker locations, *a priori* information provided a benefit for all masking patterns at −5 dB TMR and a smaller and less consistent benefit at 0 dB TMR. At −5 dB TMR, the largest benefit was observed with patterns 1–3 and at 0 dB TMR the largest benefit was for pattern 4. At both TMRs, the smallest off-masker benefit of *a priori* information was observed for masker pattern 5. At the on-masker locations, no consistent effect of *a priori* knowledge was observed.

These results were confirmed by submitting the data from Fig. 5(B) to a three-way repeated measures analysis of variance (ANOVA), with the factors of target location (on-masker vs off-masker), masker pattern (1–5) and TMR (0 and −5 dB). This ANOVA found a significant main effect of the target location and of the TMR, as well as a significant three-way interaction between the factors (Table IIA). Additional two-way ANOVAs were performed separately on the on-masker and off-masker data (Table IIB). No significant main effect or interaction was found for the on-masker data, while all main effects and interactions were significant for the off-masker data, confirming the trends shown in Fig. 5(B).

IV. DISCUSSION

A. Impact of maskers on detection and localization

One effect of presenting a speech target in the presence of four concurrent speech maskers was a reduction in the detectability of the target, as evidenced by the presence of miss trials (Table I), particularly at the poorer TMR. This was not surprising, given that concurrent speech maskers are well known to cause energetic and informational masking, both of which impede detection. We saw more misses for

TABLE II. (A) Three-way repeated measures ANOVA on the differences between the rms errors in the Mixed vs Fixed conditions (location × pattern × TMR). (B) Two two-way ANOVAs performed on the same data, but separately for the on-masker and off-masker locations (pattern × TMR).

(A) Main factor/interaction	d. f.	F	Signif.[a]
Location (on vs off)	1, 6	20.46	***
Pattern	4, 24	1.98	
TMR	1, 6	4.34	
Location × pattern	4, 24	2.11	
Location × TMR	1, 6	3.51	
Pattern × TMR	4, 24	0.89	
Location × pattern × TMR	40, 240	3.12	*

(B)		On-masker		Off-masker	
Main factor/interaction	d. f.	F	Signif.[a]	F	Signif.[a]
Pattern	4, 24	1.39		3.90	*
TMR	1, 6	0.42		9.70	*
Pattern × TMR	4, 24	0.27		3.79	*

[a]Significance levels: * $p<0.005$, ** $p<0.05$, *** $p<0.01$

on-masker locations (Fig. 3) consistent with previous studies showing greater masking for colocated stimuli whether the task is detection (Simpson *et al.*, 2006; Balakrishnan and Freyman, 2008) or intelligibility (Bronkhorst, 2000; Brungart, 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002).

In most previous studies that examined localization in the presence of maskers, disruptions to localization could largely be explained in terms of such reductions in detectability (Good and Gilkey, 1996; Good *et al.*, 1997; Lorenzi *et al.*, 1999). In the present study, however, localization performance was only measured for trials in which listeners reported to have heard the target (see also Simpson *et al.*, 2006). Using this approach, we still found that target localization was strongly degraded by the presence of the speech maskers. The maskers increased rms errors, depending on the configuration of the target and maskers and the TMR. The presence of false alarms on the catch trials (Table I) suggests that some of these effects may be a result of the listeners' tendency to respond and indicate a (probably random) target location even on trials on which they did not hear the target. However, these false alarms were very rare in the 0 dB TMR condition, and this condition gave rise to qualitatively similar patterns of results to the −5 dB TMR condition. Thus we are confident that random responses had only a minor impact on the results. Another possibility is that listeners would on occasion erroneously attribute the location of a masker to the target. It is likely that this "feature-binding" confusion would be more of a problem at the lower TMR where the target voice is more poorly segregated from the mixture and hence less distinct.

We compared the overall effect of maskers on speech localization in our study to that reported by Simpson *et al.* (2006). On average, our listeners showed rms errors of 7° in the control condition and 13° in the Mixed condition at 0 dB TMR. In a similar condition, with four different-sex maskers at 0 dB TMR, Simpson *et al.* (2006) reported rms errors (in the left-right dimension) of 8° in the control condition and 14° in the mixture case. This good correspondence in performance between our subjects and those of Simpson *et al.* (2006) suggests that the fact that our study was conducted in a reverberant office (rather than in an anechoic environment) did not increase errors in quiet or in a mixture. This was somewhat surprising given that reverberation has been shown to affect the segregation of competing speech sounds (Lavandier and Culling, 2007) as well as the localization of one sound in the presence of another (Braasch and Hartung, 2002; Kopčo *et al.*, 2007) in other studies using simpler stimuli.

B. Effects of *a priori* information on localization

Many previous studies have demonstrated detrimental effects of stimulus uncertainty on target detection and identification in complex auditory mixtures (for review, see Kidd *et al.*, 2008b). Most of these studies have examined spectral or temporal uncertainty in the target or the masker(s). A handful of studies that examined uncertainty in the spatial domain found that within- or across-trial variability in the target location can disrupt intelligibility (Kidd *et al.*, 2005; Brungart and Simpson, 2007; Best *et al.*, 2008). Variability or uncertainty in masker location, on the other hand, appears to have only a minor effect on target detection for the case of simultaneous noises (Fan *et al.*, 2008) and no effect at all on target intelligibility in the case of competing sentences (Jones and Litovsky, 2008). Furthermore, previous studies of spatial cuing for simple sound localization tasks have not found robust benefits (Spence and Driver, 1994; Sach *et al.*, 2000; Kopčo *et al.*, 2001).

The main hypothesis examined in this study was that the disruptive effect of maskers on speech localization in a multitalker mixture would be mitigated by providing the listener with *a priori* information about the masker locations. The rationale was that attentional processing may become important for localization in complex environments in which there is strong competition for processing resources. Consistent with this hypothesis, we found that the rms error was reduced by approximately 0.5° (or 15%) at 0 dB TMR and by 1° (or 20%) at −5 dB TMR.

We also predicted that the benefit of *a priori* information would be more pronounced for the off-masker target locations. In fact, we found that all of the benefit was restricted to the off-masker locations, at which the error was reduced by approximately 1° (or 31%) at 0 dB TMR and by approximately 2° (or 35%) at the −5 dB TMR. The improvement in performance for the off-masker targets is likely due to reassignment of processing resources to the off-masker locations (away from the known masker locations). However, the observed recovery from masking was not complete, showing that factors outside listener's strategic control also limited performance. These factors likely include direct acoustic interference of the competing sounds with the target sound and limitations in the binaural system's abilities to extract the relevant acoustic cues.

Another hypothesis was that the listeners would benefit more from the *a priori* information about the masker locations if the masker distribution was simple. This hypothesis was confirmed only partially. As expected, the smallest benefit was observed for pattern 5, where maskers were most distributed, thus making it difficult to strategically allocate attention to (or away from) a particular region. Interestingly, for the remaining patterns, the size of the benefit of *a priori* information depended on the TMR. At 0 dB TMR, the largest benefit was observed for pattern 4, whereas at −5 dB TMR the largest benefit was provided for patterns 1–3. While this result makes it difficult to identify the strategy employed by the listeners, it suggests that the strategy might change as the difficulty of the task increases.

C. Potential mechanisms

As a final note, it is worth considering the mechanisms that might underlie sound localization in a complex speech mixture, with a view to understanding both the impact of maskers on accuracy and the moderating influence of prior knowledge about the their spatial arrangement.

Faller and Merimaa (2004) proposed a model that showed that the robust localization of sounds in the presence of distractors reported in previous studies (e.g., Hawley

et al., 1999) could be predicted by considering binaural processing only at points in time when interaural coherence is higher than a (relatively high) critical threshold. Their model was successful at extracting independent binaural parameters for five simultaneous speech stimuli (in their case positioned at 0°, ±30°, and ±80° azimuth). This model is unlikely to handle stimuli like ours because it assumes that there are gaps in the masker profile that allow the listener to have a clear glimpse at the target. Such gaps were minimal in our stimuli because of the brief and synchronized nature of the five utterances. However, it is possible that if the threshold coherence criterion was sufficiently lowered, the model would give reasonable outputs and might even predict the increases in localization error in the presence of maskers observed here.

In order to describe the effects of masker spatial uncertainty on localization accuracy, it seems necessary to invoke more central mechanisms such as endogenous orientation (Posner and Petersen, 1990) whereby responses are modulated by attention and/or expectation. One challenge for any simple model of orienting, however, is to explain why the benefit of *a priori* information depended on the TMR and the specific masker pattern.

V. SUMMARY

Localization of a monosyllabic speech target is degraded by the presence of concurrent speech maskers, particularly for poorer TMRs, even when reduced detectability is accounted for. The impact of the concurrent maskers depends in a complex way on the target/masker configuration and whether or not the target is spatially coincident with a masker. Listeners can use *a priori* information about the location of maskers to mitigate their adverse effects on target localization, in particular, for targets that do not coincide with maskers.

ACKNOWLEDGMENTS

Work supported by grants from the Human Frontier Science Program (to N.K.), the NIH Grant No. R03 TW007640 (which partially supported N.K.), the Australian Research Council (to S.C.), and the University of Sydney Postdoctoral Research Fellowship (to V.B.).

[1]The rms errors reported here were computed relative to the mean control responses. When rms errors were defined with respect to the actual target locations, the results were similar. We chose to use the rms errors relative to the control responses since they reflect the perceived locations and account for possible inaccuracies in the placement of the target speakers and in the response measurement system.

Abouchacra, K. S., Emanuel, D. C., Blood, I. M., and Letowski, T. R. (**1998**). "Spatial perception of speech in various signal to noise ratios," Ear Hear. **19**, 298–309.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**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.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

Best, V., Ozmeral, E. J., Kopčo, N., and Shinn-Cunningham, B. G. (**2008**). "Object continuity enhances selective auditory attention," Proc. Natl. Acad. Sci. U.S.A. **105**, 13174–13178.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane: I. Psychoacoustical data," Acust. Acta Acust. **88**, 942–955.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta. Acust. Acust. **86**, 117–128.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2007**). "Cocktail party listening in a dynamic multitalker environment," Percept. Psychophys. **69**, 79–91.

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196.

Drullman, R., and Bronkhorst, A. W. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

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.

Fan, W. L., Streeter, T. M., and Durlach, N. I. (**2008**). "Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli," J. Acoust. Soc. Am. **124**, 36–39.

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., 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.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 349–376.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Jones, G. L., and Litovsky, R. Y. (**2008**). "Role of masker predictability in the cocktail party problem," J. Acoust. Soc. Am. **124**, 3818–3830.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Jr., Best, V., and Mason, C. R. (**2008a**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008b**). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Handbook of Auditory Research, New York), pp. 143–190.

Kopčo, N., Best, V., and Shinn-Cunningham, B. G. (**2007**). "Sound localization with a preceding distractor," J. Acoust. Soc. Am. **121**, 420–432.

Kopčo, N., Ler, A., and Shinn-Cunningham, B. G. (**2001**). "Effect of auditory cuing on azimuthal localization accuracy," J. Acoust. Soc. Am. **109**, 2377.

Lavandier, M., and Culling, J. F. (**2007**). "Speech segregation in rooms: Effects of reverberation on both target and interferer," J. Acoust. Soc. Am. **122**, 1713–1723.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Posner, M. I., and Petersen, S. E. (**1990**). "The attention system of the human brain," Annu. Rev. Neurosci. **13**, 25–42.

Sach, A., Hill, N., and Bailey, P. (**2000**). "Auditory spatial attention using interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **26**, 717–729.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**, 182–186.

Simpson, B. D., Brungart, D. S., Iyer, N., Gilkey, R. H., and Hamil, J. T. (**2006**). "Detection and localization of speech signals in the presence of competing speech signals," in Proceedings of the International Conference

Auditory Display, pp. 129–133.
Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free field listening II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.
Yost, W. A., Dye, R. H., Jr., and Sheft, S. (**1996**). "A simulated "cocktail party" with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston,), pp. 255–276.

Using a reflection model for modeling the dynamic feedback path of digital hearing aids[a)]

Guilin Ma[b)] and Fredrik Gran
Research Group, GN ReSound A/S, Lautrupbjerg 9, 2750 Ballerup, Denmark

Finn Jacobsen and Finn Agerkvist
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, 2800 Kgs. Lyngby, Denmark

(Received 6 May 2009; revised 16 September 2009; accepted 13 December 2009)

Feedback whistling is one of the severe problems with hearing aids, especially in dynamic situations when the users hug, pick up a telephone, etc. This paper investigates the properties of the dynamic feedback paths of digital hearing aids and proposes a model based on a reflection assumption. The model is compared with two existing models: a direct model and an initialization model, using the measured dynamic feedback paths. The comparison shows that the proposed approach is able to model the dynamic feedback paths more efficiently and accurately in terms of mean-square error and maximum stable gain. The method is also extended to dual-microphone hearing aids to assess the possibility of relating the two dynamic feedback paths through the reflection model. However, it is found that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290989]

PACS number(s): 43.66.Ts, 43.20.El, 43.60.Ac, 43.60.Uv [AJZ] Pages: 1458–1468

I. INTRODUCTION

Feedback is a severe problem in hearing aids, which limits the maximum gain that can be achieved.[1] The acoustic feedback path is the most significant contributor to the feedback signal although electrical and mechanical paths also exist.[2] A widely adopted approach to acoustic feedback suppression is feedback cancellation, where the acoustic feedback signal is estimated and then subtracted from the input signal to remove feedback.[3] The maximum stable gain (MSG) that can be obtained by feedback cancellation depends on how accurately the feedback path is estimated. A perfect match between the modeled and the real feedback path will cancel the feedback signal completely, and the system will be stable for any amount of amplification.[4]

A typical acoustic feedback path in the hearing aids includes the effects of the hearing-aid amplifier, receiver, microphone, the acoustics of the vent or leak, as well as the external acoustics. Except for the fitting and external acoustics, all the parts are nearly constant or very slowly varying with a time frame of weeks or years[5] when the user has put on the hearing aid. Assume that the user stays still in a stationary acoustic environment, and the feedback path will be almost static. This feedback path is referred to as a "static feedback path." However, in practice, the feedback path is usually subject to dynamic changes. The causes of these changes fall into two categories: (a) changes in the hearing-aid fitting, including jaw movement, smile, yawn, and cerumen; (b) changes in the external acoustics, including the factors such as room reverberation, a hand, or telephone handset near the ear. The variant feedback path due to these changes is called "dynamic feedback path." In this paper, only the dynamic feedback path for a specific position along the changing course is studied. The temporal characteristics, such as the changing rate at that position, are not addressed.

The properties of feedback paths have been investigated intensively in previous studies,[5,2,6] where the effect of vent size, the variability of feedback due to smiling, and handset proximity, were discussed. A common observation in these studies is that the largest variation in the acoustic feedback path was found when a hand or telephone handset was placed near the ear. For behind-the-ear (BTE) devices, the maximum variation can go up to 20 dB, whereas for in-the-ear (ITE) device, it can even go up to 27 dB in the 2–5 kHz frequency range compared with the static feedback path.[6]

The modeling of the feedback path has also been discussed intensively. However, the majority of the previous work has focused on the static feedback path, trying to model each static component, such as cavities and transducers, by an equivalent electroacoustic model.[7–9] For dynamic feedback path modeling, there is not very much research up to the present. One study that is relevant measured the variability of feedback path under various conditions and introduced an uncertainty bound model to calculate the robust stability condition for the hearing-aid system.[2] It provided knowledge about one aspect of the dynamic feedback path properties, i.e., the extent of variability, but did not discuss other acoustic properties and did not investigate the modeling issue either. Another relevant study dealt with feedback cancellation

[a)]Portions of this work were presented at the 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) in April 2009.

[b)]Also at Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark. Author to whom correspondence should be addressed. Electronic mail: gm@elektro.dtu.dk

 0001-4966/2010/127(3)/1458/11/$25.00

in room reverberation.[10] The focus of that study was the performance of feedback cancellation instead of the properties of the feedback path and its modeling in reverberation.

To track the dynamic aspect of feedback paths in the real world, feedback cancellation in many hearing aids today includes an adaptive feedback path model, where a simple adaptive finite-impulse-response (FIR) filter is used instead of an infinite-impulse-response (IIR) filter, since adaptive IIR filter suffers from the problem of computational complexity, instability, and local minima.[11] Thus, the estimate of dynamic feedback path is essentially a procedure of updating the coefficients of the adaptive FIR filter.

In general, two models with different ways of using the adaptive FIR filter have been proposed to model the dynamic feedback paths. In the first model, such as the systems proposed in some previous studies,[12,13] the overall feedback path is represented with the FIR filter. This model is referred to as the "direct model." The second model, such as the systems proposed in other studies,[3,14] incorporates a fixed model and an adaptive FIR filter. The fixed model represents the slowly varying portion of the feedback path (microphone, amplifier, and receiver), whereas the adaptive filter represents the rapidly varying portion (mainly the change of fitting and external acoustics). The fixed model can be either captured by an initialization,[3] referred to as an "initialization model" in this paper, or roughly approximated as a high-pass filter.[14] Compared with the direct model, the initialization model generally needs a shorter FIR filter and is computationally more efficient but requires an additional procedure in the hearing-aid fitting.

No matter which model is used, effective feedback cancellation requires that the FIR filter should adapt fast enough given the variability of the feedback paths and should be long enough to model all of the salient features.[10] The feedback path varies at different rates under different conditions. For example, jaw movements will produce changes with a time frame of seconds/minutes.[15] The picking up of a telephone handset changes the feedback path with a time frame of milliseconds/seconds. Production and disposal of cerumen will result in changes with a time frame of weeks and months, while aging causes changes with a time frame of years.[5] Usually the adaptation speed is not a problem even in the extreme cases as long as a proper adaptation scheme, for example, frequency-domain adaptation, is used. However, in practice, to maintain sound quality and minimize the artifacts resulted from adaptation errors when a spectrally colored signal is inputted,[13] the adaptation usually has to slow down and is not fast enough in some occasions. These adaptation problems, although prominent and complicated in the feedback cancellation system, are more related to the on-line adaptation scheme and the characteristics of the input signals, instead of the models of the feedback paths. The focus of this paper is on the modeling. Therefore, these adaptation issues are not addressed in this paper.

As for the sufficient length of the FIR filter in the model, the natural questions are as follows: how many orders of the FIR filter would be "enough" in various dynamic situations, and is there any other model that is more effective in modeling the dynamic feedback paths? In order to address these issues, the characteristics of the dynamic feedback path in addition to variability should be examined. This paper investigates these properties for the most adverse proximity situation that happens frequently and challenges the feedback cancellation most. This paper also describes a novel reflection based model (proposed in a previous publication[16]) in detail and extends the previous work thoroughly. The basic idea of the reflection model is to assume that the dynamic feedback path consists of multiple propagation paths and estimate the delays and gains associated with these paths directly. It differs from the two existing models in that the reflection model has an underlying physical assumption, and is more generalized by allowing fractional delays in the model, which will be described in more detail in Sec. II. The effectiveness of the new model is compared with the existing two models mentioned above, i.e., the direct model and the initialization model, using data from measurements. The comparison shows that the proposed model is superior in terms of mean-square error (MSE) and MSG.

For a hearing aid with dual microphones, due to the short distance between the two microphones (normally less than 2 cm), the two feedback paths are similar in most situations. This paper investigates the possibility of relating the two dynamic feedback paths using the initialization model and the reflection model. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths.

The outline of this paper is as follows: in Sec. II, two traditional models are explained and a new reflection model is proposed. In Sec. III, the measurement configuration and procedure are described. The properties of the dynamic feedback paths are discussed in Sec. IV. In Sec. V, the performance of the proposed model is compared with the existing two models and the possibility of using this model for relating the two dynamic feedback paths of a dual-microphone hearing aid is explored based on the measured data. Concluding remarks and directions for future work are given in Sec. VI.

II. MODELS FOR THE DYNAMIC FEEDBACK PATH

The general diagram of adaptive feedback cancellation is depicted in Fig. 1. It should be noted that the impulse responses of the microphone and receiver have been included in the impulse response of the feedback path $b(n)$. Although there are various feedback models as mentioned in Sec. I, their estimation of $b(n)$ can all be represented in a compact form $\hat{b}(n,\theta)$, where n denotes the discrete-time index and θ is a particular parameter set of a model. For example, when the feedback path is modeled by a FIR filter, θ represents the FIR coefficients. The principle of feedback cancellation is to adjust the parameters θ in the feedback model so that the modeled feedback path $\hat{b}(n,\theta)$ approximates the true feedback path $b(n)$. The output $v(n)$ is the instantaneous estimation of the feedback signal $f(n)$ and is subtracted from the input signal $s(n)$ to remove the feedback.

In principle, the impulse response $b(n)$ has infinite duration. However, the amplitude of $b(n)$ decays very fast, as

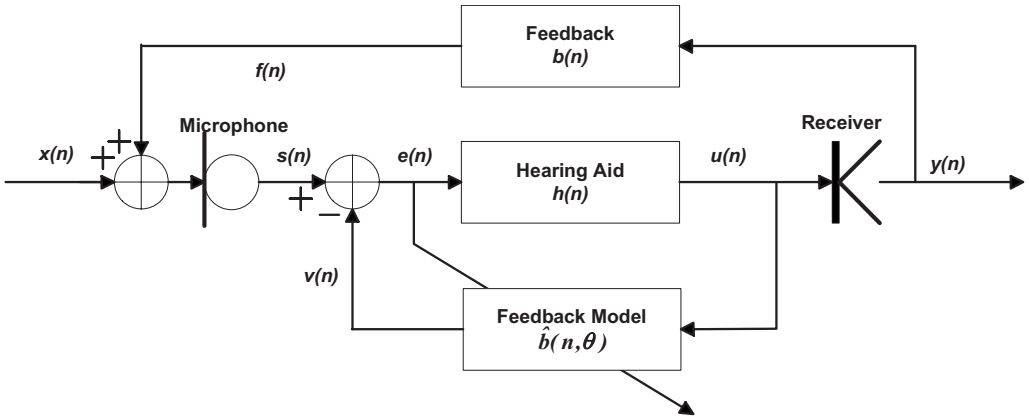


FIG. 1. General diagram of adaptive feedback cancellation. The input to the hearing-aid processing is $s(n)$, which is the sum of desired input signal $x(n)$ and the feedback signal $f(n)$. The processed hearing-aid signal is $u(n)$. The signal output into the ear canal is $y(n)$. The impulse response of the feedback path is $b(n)$, and $v(n)$ is the estimate of $f(n)$ from the modeled feedback path $\hat{b}(n,\theta)$.

shown later in Fig. 5. It is assumed in the following that both $b(n)$ and $\hat{b}(n,\theta)$ are truncated to a sufficient length so that the energy loss in the impulse response due to the truncation is at least 35 dB below the total energy of $b(n)$ or $\hat{b}(n,\theta)$.

To investigate each feedback model's capability in capturing the true feedback path, the parameters θ for each model is optimized by minimizing the difference between the modeled feedback $\hat{b}(n,\theta)$ and the actual feedback path. This minimization formulated in the frequency domain is as follows:

$$\theta_{\text{opt}} = \arg\min_{\theta} \|\mathbf{F}^H(\hat{\mathbf{b}}(\theta) - \mathbf{b})\|_2^2, \tag{1}$$

$$\mathbf{b} = [b(0), \ldots, b(L-1)]^T, \tag{2}$$

$$\hat{\mathbf{b}}(\theta) = [\hat{b}(0,\theta), \ldots, \hat{b}(L-1,\theta)]^T, \tag{3}$$

$$\mathbf{F} = [\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{L-1}] \tag{4}$$

$$\mathbf{f}_k = [1, e^{j\omega_k}, \ldots, e^{j\omega_k(L-1)}]^T, \tag{5}$$

where $\|\ \|_2$ denotes the Euclidean norm, L is the length of the truncated impulse responses in samples, $\omega_k = 2\pi k/L$, $k = 0, 1, \ldots, L-1$, $\mathbf{F}$ is the discrete Fourier transform matrix, and the superscripts T and H denote the transpose and conjugate transpose of a matrix or vector, respectively.

Based on the framework above, Secs. II A–II C will detail each feedback path model by specifying its parameter structure θ, the corresponding optimization procedure, and the analytic optimal solution if it exists. The optimal solution, either in analytic or in numerical form, represents the capability of each feedback model to represent the actual feedback path $b(n)$.

A. Existing models for dynamic feedback paths

The existing two models for dynamic feedback path are the direct model and the initialization model, both of which can be formulated as a cascade of prefiltering and an adaptive FIR filter.

Let $b_0(n)$ and $w(n)$ denote the impulse response of the prefiltering and the adaptive filter, respectively. The modeled feedback $\hat{b}(n,\theta)$ is then the convolution of $b_0(n)$ and $w(n)$, i.e.,

$$\hat{b}(n,\theta) = w(n) \odot b_0(n) = \sum_{l=0}^{M-1} w(l) b_0(n-l), \tag{6}$$

where M is the order of $w(n)$, and $\odot$ is the convolution operator.

1. Direct model

The feedback path $b(n)$ is always associated with a leading delay, which is the sum of the processing delay in $h(n)$ and the acoustic delay in the transmission path. In the "direct model," the FIR filter $w(n)$ is used to model the "active" portion of the feedback path, whereas the prefiltering $b_0(n)$ is simply a delay of D samples introduced to match the leading delay in $b(n)$ and provide a more efficient use of the limited number of taps in $w(n)$. The parameters are therefore of the form

$$b_0(n) = \begin{cases} 1, & n = D \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

$$\theta = \{w(n), D\}. \tag{8}$$

The optimal parameters $w_{\text{opt}}(n)$ and D_{opt} can be obtained by solving the optimization problem in Eqs. (1)–(8).

If the delay D is fixed, the least-squares solution is straightforward:

$$\mathbf{w}_{\text{opt}} = [b(D), \ldots, b(D+M-1)]^T. \tag{9}$$

If D is not fixed, an easy procedure to obtain the optimal solution is to search for an integer delay D that minimize the distance between $\hat{b}(n,\theta)$ and $b(n)$.

2. Initialization model

To model the feedback path accurately, the direct model in Sec. II A 1 usually needs a very high-order adaptive FIR filter $w(n)$ to cover the active range of the feedback path. One way to improve the modeling efficiency is to use the initialization model.

The feedback path $b(n)$ consists of slowly varying portions such as the responses of microphone, amplifier, receiver, etc. The idea of the initialization model is to use a fixed model, which is initialized during an additional off-line initialization procedure in the hearing-aid fitting, to model these slowly varying portions, and use a short adaptive FIR filter to represent the rapidly varying portions (mainly the change in fitting and external acoustics).

In such an initialization procedure, a "nominal feedback path" is first measured, which is defined as the feedback path without any external reflectors or enclosures near the hearing aid. When the hearing aid is put into use in daily life, the adaptive filter $w(n)$ only needs to model the part that is different from the nominal initialization to capture the time-varying dynamic feedback path. Since the impulse responses of microphone, receiver, etc., are nearly invariant from the initialization to dynamic situations, this different part is mainly resulted from the change in the hearing-aid fitting and external acoustic environments,[2] which can be modeled by an adaptive filter with a lower order.

In the initialization model, the prefiltering $b_0(n)$ in Eq. (6) is a fixed model to represent the impulse response of the nominal feedback path obtained in the initialization. In practice, it can be realized by an autoregressive moving average model (ARMA) model.[3] To avoid the complicated issue of how to model the initialized feedback path with an ARMA model to get the best generalized performance in feedback cancellation,[3,17] it is simply assumed that $b_0(n)$ is exactly the impulse response of the nominal feedback path. Therefore, the parameter for initialization model is of the form

$$\theta = \{w(n)\}. \tag{10}$$

Assume that the filter length of $w(n)$ is M, and the impulse response $b_0(n)$ is truncated to $L-M+1$ samples and padded with $M-1$ zeros so that the length of $\hat{b}(n,\theta)$, which is the convolution of $w(n)$ and $b_0(n)$, equals L. The impulse response of the dynamic feedback path $b(n)$ is also assumed to be of length L.

The optimal parameter $w_{\mathrm{opt}}(n)$ for real valued $w(n)$ with length M in the initialization model can be found by solving the least-squares problem with Eqs. (1)–(6) and (10):

$$\mathbf{w}_{\mathrm{opt}} = (\mathrm{diag}(\mathbf{F}^H\tilde{\mathbf{b}}_0)\tilde{\mathbf{F}}^H)^+(\mathbf{F}^H\mathbf{b}), \tag{11}$$

$$\tilde{\mathbf{b}}_0 = [\mathbf{b}_0^T, \mathbf{0}_{1\times(M-1)}]^T, \tag{12}$$

$$\mathbf{b}_0 = [b_0(0), \ldots, b_0(L-M)]^T, \tag{13}$$

$$\tilde{\mathbf{F}} = [\tilde{\mathbf{f}}_0, \tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_{L-1}], \tag{14}$$

$$\tilde{\mathbf{f}}_k = [1, e^{j\omega_k}, \ldots, e^{j\omega_k(M-1)}]^T, \tag{15}$$

where $\omega_k = 2\pi k/L$, $k=0,1,\ldots,L-1$, $\mathrm{diag}(\cdot)$ forms a diagonal matrix with diagonal elements specified in $(\cdot)$, $(\cdot)^+$ is a pseudoinverse defined as $(\cdot)^+ = ((\cdot)^H(\cdot))^{-1}(\cdot)^H$, and $\mathbf{0}_{1\times(M-1)}$ represents a zero vector of size $1\times(M-1)$. The matrix $\tilde{\mathbf{F}}$ is a partial discrete Fourier transform matrix which results from the fact that $w(n)$ is only of the length M.

B. Reflection model for dynamic feedback paths

1. Model formulation

The initialization model $\hat{b}(n,w(n))$ formulated in Eq. (6) can also be regarded as a weighted sum of the nominal impulse response $b_0(n)$ and its delayed replicas with integer delays.

In typical dynamic situations where there is room reverberation or a handset placed close to the hearing aid, the feedback path is a composition of reflection and refraction, where reflection usually dominates. In addition to attenuation or amplification in the amplitude, a certain delay is associated with each reflected component. These physical delays, when transformed into the digital domain, are possibly fractional in terms of samples. Therefore, the model in Eq. (6) is generalized to a new model as follows:

$$\hat{b}(n,\theta) = \sum_{l=0}^{M-1} w(l) b_0(n-d_l), \tag{16}$$

$$\theta = \{w(l), d_l\}, \tag{17}$$

where d_l is the delay of the lth replica, $d_l > d_{l-1} \geq 0$, and $l = 1,\ldots,M-1$. Although the delays d_l in the model equations (16) and (17) are allowed to be fractional, the actual subsample implementation needs time-domain interpolation or frequency-domain multiplication. The latter is more efficient and is given in Sec. II B 2.

In this model, the impulse response of the nominal feedback path, measured during initialization without significant disturbances such as reflections, is regarded as an approximation of the direct path. The dynamic feedback path is modeled as a sum of reflection components with delay d_l and gain $w(l)$. This model is thus named "reflection model." When $d_l = l$, the reflection model reduces to the initialization model. Therefore, the reflection model is more general and expected to capture the dynamic feedback path better than the initialization model since it represents more accurately what happens in the physical world.

2. Delay estimation

The optimal delays $d_{l,\mathrm{opt}}$ and weights $w_{\mathrm{opt}}(l)$ for the reflection model can be found by solving the optimization problem given by Eqs. (1)–(5), (16), and (17), which is a nonlinear optimization problem.

However, efficient time delay (TDE) and amplitude estimation techniques have been investigated intensively in many fields such as radar, sonar, radio navigation, geophysical/seismic exploration, wireless communication, and medical imaging. The most well-known approach is

based on the matched filter,[18] whose resolution capability is unfortunately limited to the reciprocal of the signal bandwidth. For dynamic feedback path modeling, the difference between delays can be very small due to the handset proximity effect. Therefore, TDE techniques with high resolution should be used. Among these techniques, algorithms such as multiple signal classification, linear prediction, and maximum likelihood are not very well suited here since they are best for complex-valued signals with special spectral shapes.[19] A weighted Fourier transformation and relaxation (WRELAX) based method[20] is found to be very robust to address the problem. First, the cost function is defined as follows:

$$C_1(\{w(l),d_l\}_{l=0}^{M-1}) = \left\| \mathbf{B} - \sum_{l=0}^{M-1} w(l)[\mathbf{Sa}(d_l)] \right\|_2^2, \tag{18}$$

where

$$\mathbf{B} = \mathbf{F}^H \mathbf{b}, \tag{19}$$

$$\mathbf{S} = \mathrm{diag}(\mathbf{F}^H \tilde{\mathbf{b}}_0), \tag{20}$$

$$\mathbf{a}(d_l) = [1, e^{-j(2\pi/L)d_l}, \ldots, e^{-j(2\pi/L)d_l(L-1)}]^T. \tag{21}$$

Equation (18) is simply a formulation of the inner part of Eq. (1) with Eqs. (16) and (17) inserted. Minimizing the cost function $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ with respect to $w(l)$ and d_l is the problem of interest. To address this optimization problem, denote

$$\mathbf{B}_l = \mathbf{B} - \sum_{i=0,i\neq l}^{M-1} \hat{w}(i)[\mathbf{Sa}(\hat{d}_i)], \tag{22}$$

$$C_2(w(l),d_l) = \|\mathbf{B}_l - w(l)\mathbf{Sa}(d_l)\|_2^2. \tag{23}$$

The cost function $C_2(w(l),d_l)$ assumes that the other reflection components have been estimated as $\hat{w}(i)$ and $\hat{d}_i$, where $i=0,1,\ldots,M-1, i\neq l$, except the lth component, and minimizing this cost function yields the estimates $\hat{w}(l)$ of $w(l)$ and $\hat{d}_l$ of d_l as[20]

$$\hat{d}_l = \arg\max_{d_l} |\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)|^2, \tag{24}$$

$$\hat{w}(l) = \left. \frac{\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)}{\|\mathbf{S}\|_2^2} \right|_{d_l=\hat{d}_l}, \tag{25}$$

where the asterisk denotes the complex conjugate. If $w(l)_{l=0}^{M-1}$ are assumed to be real numbers, minimizing $C_2(w(l),d_l)$ with respect to $w(l)$ and d_l yields[20]

$$\hat{d}_l = \arg\max_{d_l} \{\Re[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]\}^2, \tag{26}$$

$$\hat{w}(l) = \left. \frac{\Re[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]}{\|\mathbf{S}\|_2^2} \right|_{d_l=\hat{d}_l}, \tag{27}$$

where $\Re[\cdot]$ denotes the real part of $(\cdot)$.

To estimate $\hat{d}_l$ using Eq. (26), an initial estimate of d_l is first obtained by locating the maximum peak in the magnitude of the K-point ($K \geq L$) fast Fourier transform (FFT) of $\mathbf{S}^*\mathbf{B}_l$ with padded zeros. This process can also be regarded as a matched filtering in the frequency domain. Then a search for the $\hat{d}_l$ that maximizes $\Re\{[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]\}^2$ is made in a small range around the initial estimate based on golden section search and parabolic interpolation.[21,22]

The WRELAX estimates the delays d_l and amplitudes w_l in the frequency domain to avoid fractional delay interpolation. The general idea is to cross-correlate $b(n)$ and $b_0(n)$ in the frequency domain to find the coarse delays and gains of the replicas by identifying the peaks of the cross-correlation. Later an iterative search is performed by keeping one replica of $b_0(n)$ at a time (removing the other identified replicas from $b(n)$), repeating the cross-correlation and locating the peak to find a better delay and gain estimation for that replica. This process is iterated until the relative change in the cost function in Eq. (18) is below the threshold ϵ. The detailed steps are given below.

Step 1: For $M=1$ obtain $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}$ by using Eqs. (26) and (27).

Step 2: For $M=2$ compute $\mathbf{B}_1$ with Eq. (22) by using $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ obtained in Step 1. Obtain $\{\hat{w}(l),\hat{d}_l\}_{l=1}$ from $\mathbf{B}_1$ by using Eqs. (26) and (27). Next recompute $\mathbf{B}_0$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=1}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}_0$. Iterate these two substeps until the relative change in $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ between two consecutive iterations is below the threshold ϵ.

Step 3: For $M=3$ compute $\mathbf{B}_2$ with Eq. (22) by using $\{\hat{w}(l),\hat{d}_l\}_{l=0}^{1}$ obtained in Step 2. Obtain $\{\hat{w}(l),\hat{d}_l\}_{l=2}$ from $\mathbf{B}_2$. Next recompute $\mathbf{B}_0$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=1}^{2}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}_0$. Then compute $\mathbf{B}_1$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=0,2}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=1}$. Iterate these three substeps until the relative change in $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ between two consecutive iterations is below the threshold ϵ.

Remaining Steps: Continue similarly until M is equal to the desired number of paths.

In contrast to some estimation algorithms, such as the expectation-maximization algorithm,[23] the WRELAX is not sensitive to the initial parameters since it uses a matched filter to find these initial values. It is bound to converge to at least a local minimum point, which is at least a better solution than what the traditional matched filter can yield. The convergence speed of WRELAX depends on the time delay spacing of the different reflected paths.[20]

C. Models for dual-microphone hearing aids

For hearing aids with dual microphones, the feedback problem involves a dynamic feedback path for the front microphone, denoted as $b_f(n)$, and a path for the rear microphone, denoted as $b_r(n)$. One way of dealing with the two paths is to model them individually by using one of the three models described above. An alternative approach involves

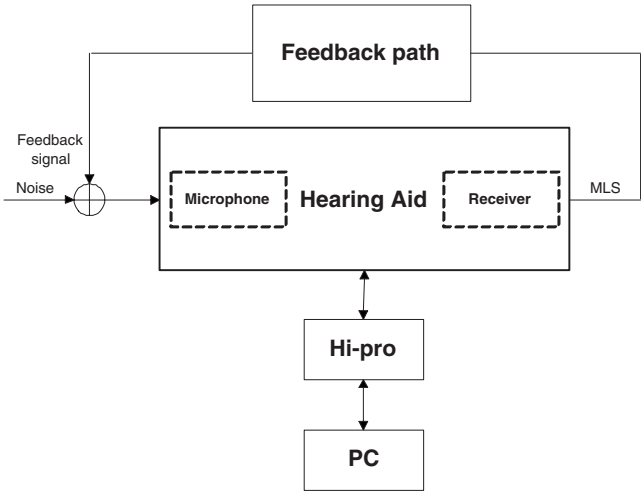


FIG. 2. The block diagram of the measurement system for measuring the feedback paths. The feedback path is illustrated which includes the external acoustics, for example, a Kemar head. The receiver sends out a MLS and the feedback signal is picked up by the microphone with ambient noise together. The data are fed into the computer through HI-PRO.

exploiting the relation between the two similar feedback paths by fitting one dynamic feedback path with the other dynamic path. There are two ways of fitting.

The first approach for the fitting, which is similar to the initialization model [cf. Eq. (6)], is

$$\hat{b}_r(n) = \sum_{l=0}^{M-1} w(l) b_f(n-l). \tag{28}$$

The second approach for the fitting, which is similar to the reflection model [cf. Eq. (16)], is

$$\hat{b}_r(n) = \sum_{l=0}^{M-1} w(l) b_f(n-d_l). \tag{29}$$

The estimate of $w_{\text{opt}}(n)$ and/or delays $d_{l,\text{opt}}$ can be found in similar ways described in Secs. II A 2 and II B.

III. MEASUREMENT OF DYNAMIC FEEDBACK PATHS

The nominal and dynamic feedback paths have been measured using a commercial open-fitting BTE device with two microphones from GN ReSound A/S. The hearing aid was mounted on the head of Kemar Manikin type 45BA made by G.R.A.S. Sound & Vibration A/S.[24]

The impulse responses of the feedback paths were measured by sending out a maximum-length sequence (MLS) with a period of 255 samples through the receiver, repeating it 1000 periods, and averaging the responses to get a high signal-to-noise ratio (SNR) for the feedback path response relative to random ambient noise. The sampling frequency was 15625 Hz. Figure 2 illustrates the measurement diagram where a HI-PRO universal programming interface for hearing instruments made by GN Otometrics A/S (Ref. 25) is used to feed the collected data into a computer to calculate the impulse responses. The detailed procedure of the impulse response measurement can also be found in Ref. 10.

The MLS is emitted at around 85 dB sound pressure level (SPL) and transmitted to the microphone through the

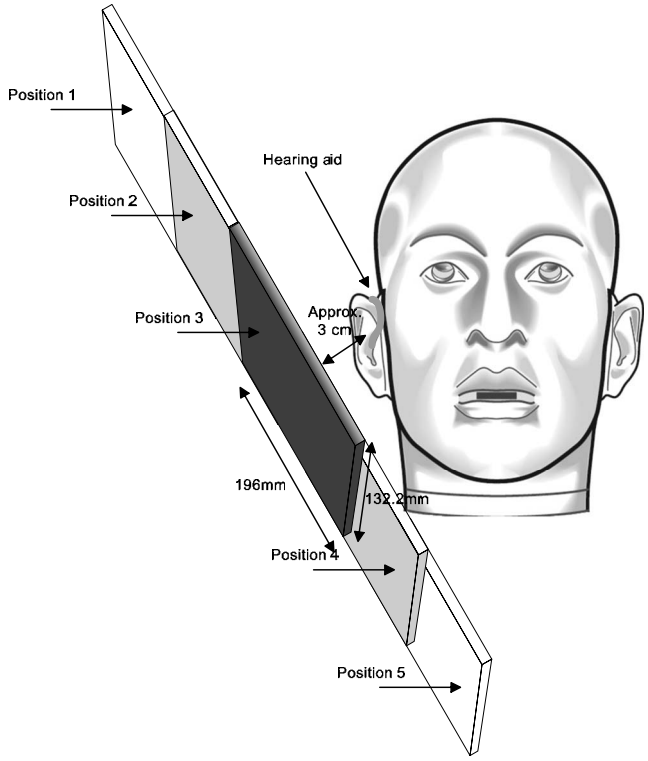


FIG. 3. The measurement setup for dynamic feedback paths with rigid surface. The hearing aid is mounted on the Kemar head. Positions 1 and 2 are symmetrical to positions 4 and 5 with regard to position 3, which is right at the side of the hearing aid.

feedback. The feedback path shapes the signal based on its frequency response. The internal noise in the microphone is less than 28 dB SPL and the ambient noise is less than 30 dB SPL. The average over 1000 cycles of MLS increases the overall SNR by 30 dB. This setup has been found to be very reliable. The preliminary test shows that when the measurements are repeated for ten times for a single feedback path, the variance in the measured impulse responses is around 36 dB below the average response, which is sufficient for feedback path modeling.

The nominal feedback paths for both the front and rear microphones are measured. The measured nominal feedback path for the front microphone is denoted as $b_{f,0}(n)$ and for the rear microphone $b_{r,0}(n)$.

Since it was found that the movement of jaw, mouth, etc., had little effect on the feedback path variations,[5,2] and the adaptive feedback cancellation in hearing aids today usually handles these situations very well in practice, the measurement of dynamic feedback paths focused on the proximity effect.

Seven dynamic feedback paths were measured for both the front and rear microphones. A rigid metal surface of dimension 196×132.6 mm^2 was used and faced to the hearing aid mounted on the Kemar head. The surface was moved along the lateral side gradually toward the hearing aid from position 1 to position 3 and later from position 3 to position 5, as shown in Fig. 3. The perpendicular distance between the rigid surface and the hearing aid was kept at around 3 cm during the movement. The impulse responses of the dynamic feedback path at these five positions measured from the front

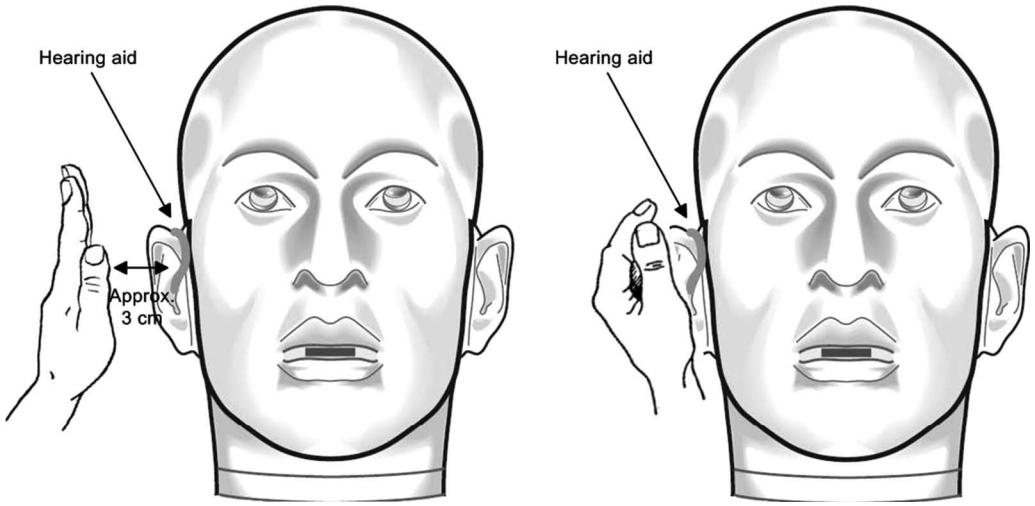


FIG. 4. The measurement setup for dynamic feedback paths with palm. The left figure represents an outstretched palm facing the hearing aid, and the right figure depicts the cupped hand around the hearing aid.

microphone are denoted as $b_{f,1}(n)-b_{f,5}(n)$ and from the rear microphone $b_{r,1}(n)-b_{r,5}(n)$. Apart from the rigid surface, two additional dynamic feedback paths were measured with an outstretched palm facing the hearing aid on its lateral side at a distance of around 3 cm $[b_{f,6}(n),b_{r,6}(n)]$ and with a palm wrapping around the hearing aid $[b_{f,7}(n),b_{r,7}(n)]$, as illustrated in Fig. 4.

Therefore, altogether eight impulse responses were measured for both front and rear microphones including the static feedback path (nominal feedback) and seven dynamic feedback paths. All the impulse responses are of the length 255 samples. The rigid surface measurements are artificial tests for reflections whereas the palm measurements mimic the most adverse situations for feedback cancellation in real life.

The frequency responses of these feedback paths are shown in Fig. 5, where only the responses from the front microphone are illustrated since the responses from the rear are similar.

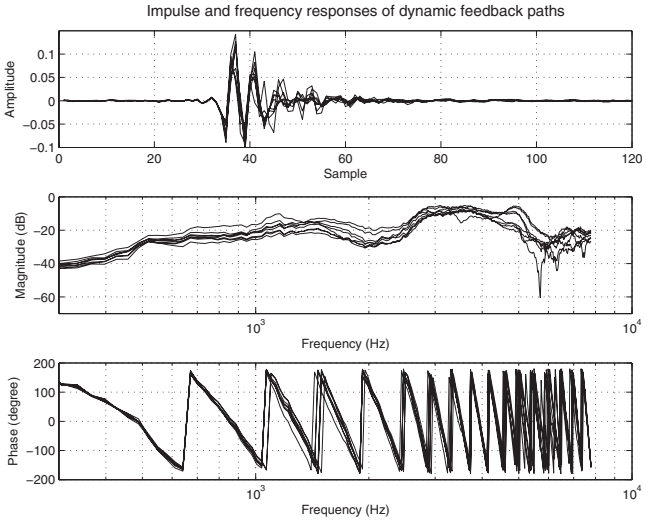


FIG. 5. Impulse responses and frequency responses of measured eight feedback paths (one nominal and seven dynamic feedback paths).

IV. PROPERTIES OF DYNAMIC FEEDBACK PATHS

In this section, the general characteristics of the measured dynamic feedback paths are investigated first. Later, the reflection model proposed in Sec. II B is used to model these paths and reveal other aspects of the characteristics of dynamic feedback paths.

A. Variability of dynamic feedback paths

The sample mean and sample estimate of the standard deviation of the measured dynamic feedback paths have been calculated as a function of frequency. The mean and standard deviation do not reveal all the properties of the dynamic feedback paths but give an indication of the dynamic range of the data. If it can be assumed that the variation in spectrum magnitude for each frequency is normally distributed, then the 95% confidence interval of the expected value for seven measurements is $[\bar{\mathbf{B}}_f-0.925\mathbf{S}_f,\bar{\mathbf{B}}_f+0.925\mathbf{S}_f]$ and $[\bar{\mathbf{B}}_r-0.925\mathbf{S}_r,\bar{\mathbf{B}}_r+0.925\mathbf{S}_r]$ for the front- and rear-microphone dynamic paths, respectively, where $\bar{\mathbf{B}}_f$ and $\bar{\mathbf{B}}_r$ are sample means, and $\mathbf{S}_f$ and $\mathbf{S}_r$ are sample estimates of the standard deviation.

They are calculated as follows:

$$\bar{\mathbf{B}}_f=\frac{1}{7}\sum_{l=1}^{7}|\mathbf{F}^H\mathbf{b}_{f,l}|, \tag{30}$$

$$\mathbf{S}_f=\sqrt{\frac{1}{7-1}\sum_{l=1}^{7}(|\mathbf{F}^H\mathbf{b}_{f,l}|-\bar{\mathbf{B}}_f)^2}, \tag{31}$$

where

$$\mathbf{b}_{f,l}=[b_{f,l}(0),\ldots,b_{f,l}(L-1)]^T. \tag{32}$$

Note that $\bar{\mathbf{B}}_f$ and $\mathbf{S}_f$ are both vectors. The symbol $|\cdot|$ in Eq. (30) and $(\cdot)^2$ in Eq. (31) denote the elementwise absolute operation and elementwise square of the vector in $(\cdot)$, respectively. The $\bar{\mathbf{B}}_r$ and $\mathbf{S}_r$ can be calculated similarly.

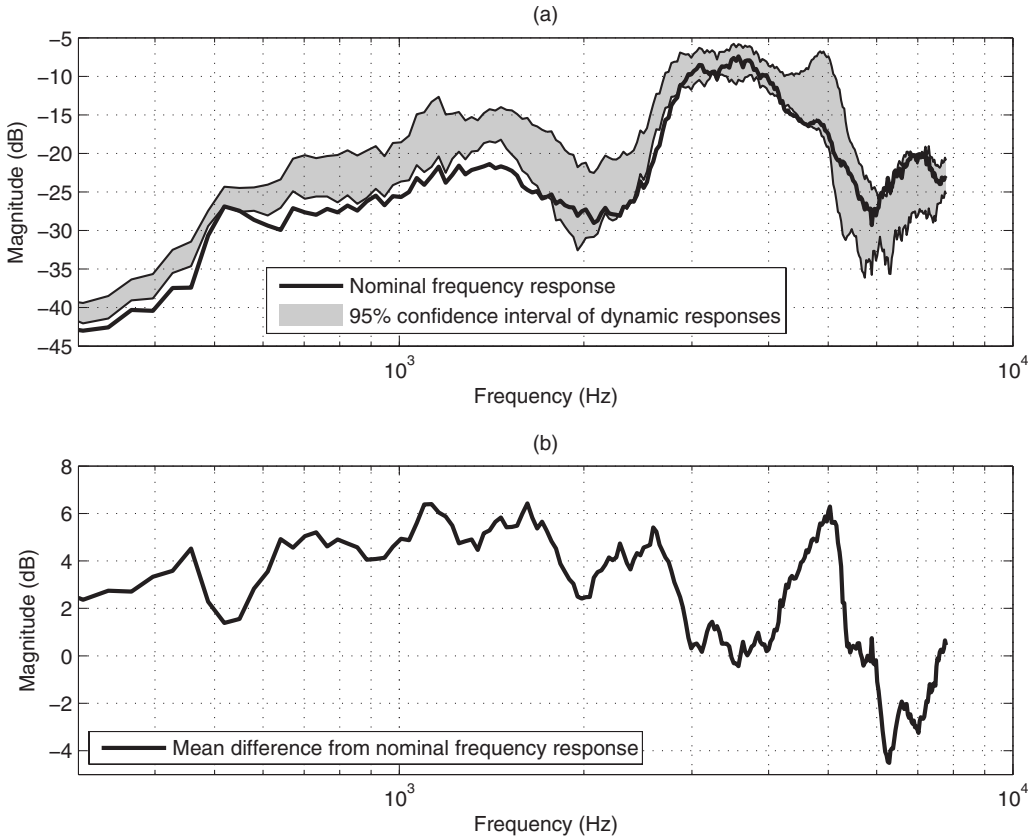


FIG. 6. Variability of dynamic feedback paths: (a) shows the nominal frequency response and a shaded region representing the 95% confidence interval of the expected frequency response of dynamic feedback paths; (b) illustrates the mean difference between the responses of dynamic feedback paths and the nominal feedback path.

The variability results are shown in Fig. 6(a). The shaded region represents the uncertainty in the expected magnitude of frequency responses of the dynamic feedback paths. It should be noted that the variability illustrated here is obtained with one single head and does not include the intersubject variability. The general trend is that at low frequencies, the uncertainty is small, but above 1 kHz, the uncertainty increases and goes up to 10 dB at some frequencies. The magnitude curve of the frequency response of the nominal feedback path lies below this shaded region at frequencies lower than 1750 Hz, which means that the rigid surface or hand in the measurements generally boosted the feedback responses at low frequencies. Above 1750 Hz, most of the nominal curve lies within the shaded region, and above 5 kHz, nearly all of the curve lies inside the region. This indicates that the change in feedback path at high frequencies is not a general boost but resulted from complicated effects such as peaks and valleys. This trend agrees with the results from other studies,[5,2,6] despite the minor differences in the measured feedback paths due to different devices, test environments, etc.

The difference in dB between the sample mean of the magnitude of dynamic feedback paths $\bar{\mathbf{B}}_f$ and the magnitude of the nominal feedback path is illustrated in Fig. 6(b), which essentially removes the effects of microphone, receiver, etc. A dominant peak was found at 5035 Hz. This can be interpreted as a strong reflection from an object placed at a distance of $1/5035 \times 343/2$ m ≈ 0.034 m, where 343 m/s is the speed of sound at 20 °C and the factor of 2 is used to calculate a single-way transmission length. This is approximately the distance between the hand/rigid surface and the hearing aid in the measurement.

B. Reflections in dynamic feedback paths

The reflection phenomenon in dynamic feedback paths with a telephone handset near the hearing aid has been pointed out in several previous papers.[3,5,6] It features peaks or other complicated effects in the high frequency range of the measured feedback responses. The dominant peak in the Fig. 6(b) also indicates the occurrence of a reflection.

However, physically, there could be many other reflections besides the dominant one, especially when the object is placed very close to the hearing aid. The reflection model proposed in this paper can be used to estimate these reflections.

In WRELAX of the reflection model, one zero is padded to all of the impulse responses of the measured feedback paths so that the truncated length L becomes 256. The number of points in the FFT for initial delay estimation, K, is set to 4096 to get a good initial guess, the search region around the initial estimate $\hat{d}_{l,\mathrm{ini}}$ in samples is set to be $[\hat{d}_{l,\mathrm{ini}} - \frac{1}{32}, \hat{d}_{l,\mathrm{ini}} + \frac{1}{32}]$, and the threshold for the stopping criterion ϵ is set as 10^{-4}.

The reflection model estimates the $w(l)$ and d_l, where $l = 0, 1, \ldots, M-1$, for all the measured data. In Figs.

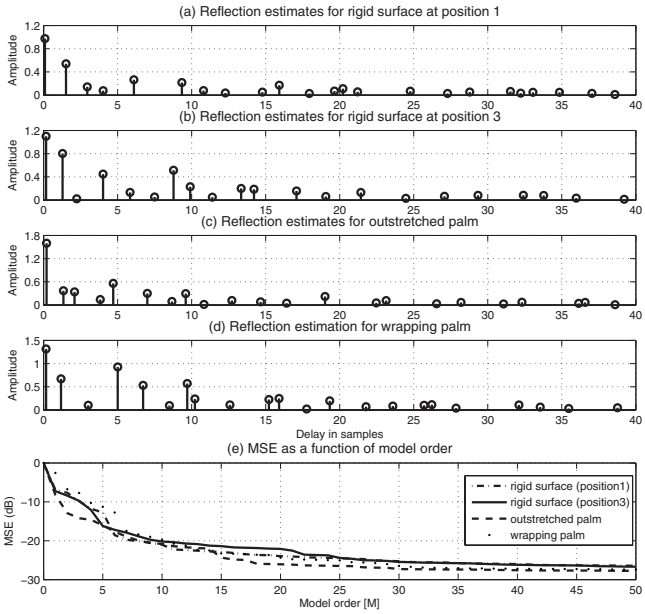


FIG. 7. (a)–(d) illustrate the reflection estimates in terms of delay in samples $\hat{d}_l$ and amplitude $\hat{w}(l)$ based on the measured data for rigid surface at positions 1 and 3, outstretched palm, and wrapping palm; (e) shows how the MSE, i.e., $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ decreases as the number of reflection components M increases.

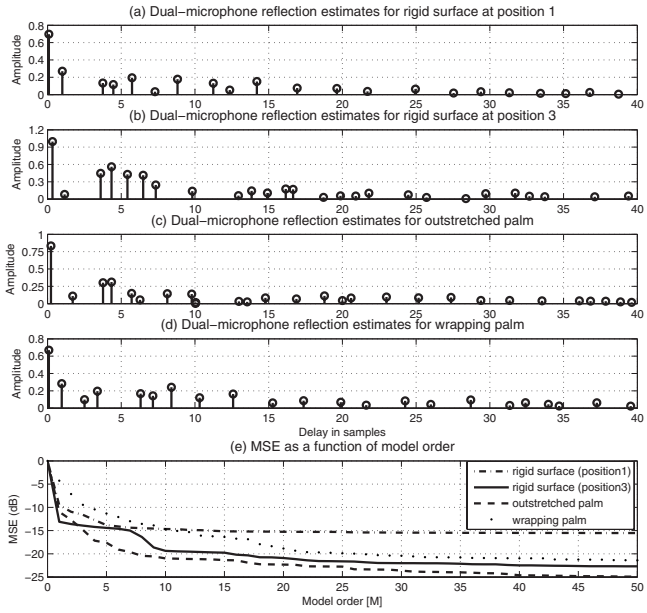


FIG. 8. (a)–(d) illustrate the reflection estimates for the dual-microphone BTE hearing aid assuming that the rear response is a sum of reflected replicas of the front response in terms of delay in samples $\hat{d}_l$ and amplitude $\hat{w}(l)$ based on the measured data for rigid surface at positions 1 and 3, outstretched palm, and wrapping palm; (e) shows how the MSE, i.e., $C_1(\{w(l),d_l\}_{L=0}^{M-1})$ decreases as the number of reflection components M increases.

7(a)–7(d), the results for the rigid surface at positions 1 and 3, and palm outstretched and wrapped around the hearing aid, are illustrated. The MSE of the estimated feedback path, i.e., $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ defined in Eq. (18), is given in Fig. 7(e) as a function of the model order. As can be seen from the figure, in all the four cases, the highest peak is located near 0 sample delay, which implies that the assumption that the nominal feedback path is approximately the direct path is valid. The reason why the amplitude is not exactly unity and the delay is not exactly 0 is that the nominal feedback path itself consists of a number of small reflections due to the presence of pinna and head.

As the delay increases, the estimated amplitude generally decays, which complies with the physical law that longer transmission path results in larger attenuation (transmission loss).[26] Almost all of the significant peaks in the four cases happen before a delay of ten samples, corresponding to a single-way transmission length of 0.106 m, which is a reasonable range for late reflection paths, e.g., a path consisting of multiple reflections between reflecting objects, the ear, and the hearing aid before it reaches the microphone.

There are fewer significant reflections (with amplitude larger than 0.4) in the cases of the rigid surface at position 1 and outstretched palm than in the cases of the rigid surface at position 3 and wrapping hand. This agrees with the fact that the reflections in the latter two cases are more complicated.

The case with the wrapping hand has the most irregular reflection pattern, in which large reflections are distributed widely in the delay range from zero to ten samples. The reflections for this situation after a delay of ten samples are generally as small as in the other cases. This shows that the wrapping hand introduces large reflections with short delays.

In the experiments with the rigid surface and outstretched palm, the distance was kept at around 3 cm, corresponding to a delay of 2.73 samples and a peak at 5717 Hz in the magnitude response. The peak has been verified in Sec. IV A, whereas for all the cases in Figs. 7(a)–7(d), the peak does not seem to be very dominant, although there is a reflection component located at around 2.73 samples. A possible reason is that the dominant reflections were influenced by the presence of the pinna and the head, depending on how the reflecting object was positioned. However, this influence is different for different measurements. Therefore, it is averaged out and the dominant peak shows up at a correct location in the average magnitude response [see Fig. 6(b) at 5035 Hz].

For the dual-microphone situation, the estimation results are shown in Fig. 8. For each case, the response of the feedback path at the rear microphone is fitted by that at the front microphone using the proposed reflection model. As seen from Fig. 8(e), the MSE converges at a relatively higher level compared with Fig. 7(e), especially for the case where the rigid surface is placed at position 1. Therefore, the usage of a reflection model to relate the two feedback paths of a dual-microphone BTE hearing aid is not very effective.

V. PERFORMANCE OF THE PROPOSED MODEL

In this section, the proposed reflection model is compared with the two existing model in terms of MSE and MSG. The performance of the reflection model for dual-microphone situation is also investigated.

A. Performance metric of feedback path models

To evaluate the performance of a feedback model, two metrics are usually used: MSE and MSG. The MSE de-

scribes the average distance between the modeled and the real feedback paths. The MSG indicates the maximum gain without instability assuming a flat frequency response in the hearing-aid processing and the worst case for the phase. The MSG is determined by the frequency at which the mismatch between the feedback model and the actual feedback path is the largest.[10] Assume that all the models are optimized, the resultant MSE and MSG are defined as MSE_c and MSG_c, respectively,

$$\mathrm{MSE}_c = 20\log_{10}\left(\frac{\|\mathbf{F}^H\hat{\mathbf{b}}(\theta_{opt}) - \mathbf{b}\|_2^2}{\|\mathbf{F}^H\mathbf{b}\|_2^2}\right), \tag{33}$$

$$\mathrm{MSG}_c = 20\log_{10}\left(\min_k \frac{1}{|\mathbf{f}_k^H(\hat{\mathbf{b}}(\theta_{opt}) - \mathbf{b})|}\right). \tag{34}$$

With a specific model and parameters θ, the MSE_c and MSG_c are the lowest achievable MSE and the highest achievable MSG, respectively. MSE_c and MSG_c are, in fact, limited by the amount of undermodeled feedback path, the residual feedback path that cannot be modeled due to the limited number of degrees of freedom in the parameter θ and/or the lack of flexibility in the model form. A more descriptive model with larger degrees of freedom in the parameters θ will yield less undermodeling, lower MSE_c and larger MSG_c.

B. Results

For each measured dynamic feedback path, the parameters in the models were first optimized to calculate the MSE_c and MSG_c. The filter length M was varied from 1 to 50. To achieve the best performance, the delay D was not fixed in the direct model.

It is found that for all the seven dynamic paths and all the values M, the reflection model outperforms the initialization model and the direct model in terms of MSE_c and MSG_c. The direct model performs worst in all the cases. To demonstrate the performance of each model in dynamic situations, MSE_c and MSG_c are averaged over the seven dynamic paths. The results are illustrated in Figs. 9 and 10. The results for the dual-microphone models are also included.

As seen in the figures, the general trend of MSE_c and MSG_c is similar, the larger M the smaller MSE_c and the larger MSG_c. The slight difference in the pattern between MSE_c and MSG_c is because the optimization of parameters in the models is based on MSE instead of MSG directly.

The reflection model is superior to the other two models, especially when M is between 6 and 21. In practice, M is usually chosen between 10 and 20 to assure a fast convergence. In this region, the reflection model yields 5–6 dB higher MSE_c, 5–7 dB higher MSG_c than the initialization model, and 8–10 dB higher MSE_c, and 9–11 dB higher MSG_c than the direct model. To achieve a 25 dB MSG_c, the direct model needs 31 orders and the initialization model needs 16 orders, whereas the reflection model only needs seven reflection replicas of the nominal impulse response. Moreover, including 11 reflection replicas in the reflection model yields around the same MSG_c as the initialization model with 50-order FIR filter. This shows that 11 reflections are enough to model the dominant characteristics of the dynamic feedback paths resulted from a hand or rigid surface placed close to the hearing aid.

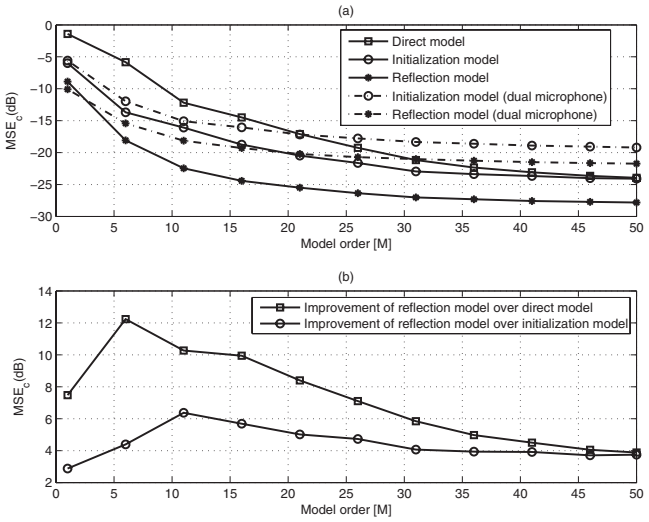


FIG. 9. Comparison of the models for dynamic feedback path modeling in terms of MSE_c. (b) shows the MSE_c improvement of the reflection model over direct model and initialization mode. The horizontal axis is the filter length in the direct model and initialization model or the number of reflection components in reflection model.

It should be noted that the reflection model does not yield significant improvement when used in the dual-microphone case. This is because in a complex acoustic environment, the relation between the two feedback paths is very complicated and even more difficult to model than the feedback paths themselves.

VI. CONCLUSIONS AND FUTURE WORK

This paper describes a novel reflection model for the dynamic feedback path in digital hearing aids. The reflection

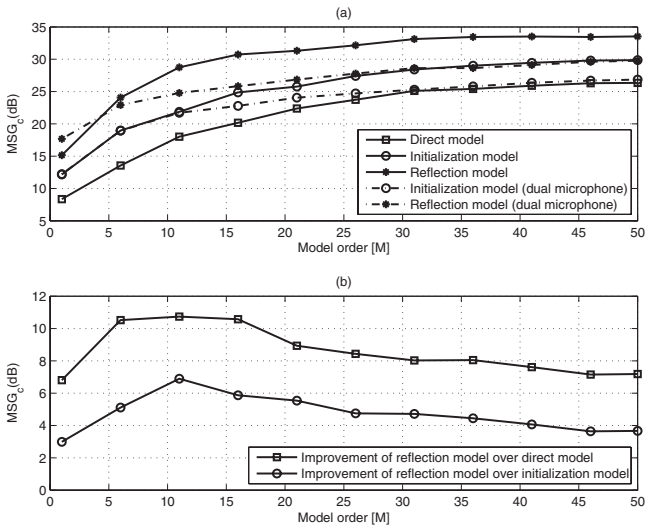


FIG. 10. Comparison of the models for dynamic feedback path modeling in terms of MSG_c. (b) shows the MSG_c improvement of the reflection model over direct model and initialization mode. The horizontal axis is the filter length in the direct model and initialization model or the number of reflection components in reflection model.

properties in addition to the variability of feedback path, when the objects such as a palm or handset are placed close to the hearing aid, have been investigated.

The results based on the measurement of a commercial hearing aid show that the proposed model has a better ability of capturing the dynamic feedback path in these difficult real-life situations and is superior to the existing two models in terms of MSE and MSG. To achieve the same level of MSE or MSG, the number of components required in the proposed model is fewer than the orders of the FIR filter required in the two traditional models.

The results also show the minimum order of the adaptive filter in the two existing models to achieve a certain MSG in the dynamic situations, which could serve as a useful indication in practice for choosing the order of the adaptive filter in feedback cancellation. For the reflection model, 11 reflection replicas are enough to model the dominant characteristics of the proximity effect in dynamic feedback paths.

This paper also investigates the possibility of relating the two feedback paths of a dual-microphone hearing aid for modeling the dynamic feedback paths. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths.

The future work is to develop an on-line adaptation scheme for this reflection model so that it can improve the performance of feedback cancellation in adverse dynamic situations. The first step to this adaptation scheme is to find a more efficient way of estimating the fractional delays. The frequency-domain approach will be preferred since the implementation of a fractional delay in the time domain requires interpolation. The next step is to find a proper frequency-domain adaptive filtering approach to estimate the dynamic feedback paths based on the estimated delays.

[1]S. F. Lybarger, "Acoustic feedback control," in *The Vanderbilt Hearing-Aid Report*, edited by Studebaker and Bess (Monographs in Contemporary Audiology, Upper Darby, PA, 1982), pp.87–90.

[2]B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," J. Acoust. Soc. Am. **107**, 2665–2673 (2000).

[3]J. Kates, "Constrained adaptation for feedback cancellation in hearing aids," J. Acoust. Soc. Am. **106**, 1010–1019 (1999).

[4]J. Kates, *Digital Hearing Aids* (Plural, San Diego, CA, 2008).

[5]J. Hellgren, T. Lunner, and S. Arlinger, "Variations in the feedback of hearing aids," J. Acoust. Soc. Am. **106**, 2821–2833 (1999).

[6]M. Stinson and G. Daigle, "Effect of handset proximity on hearing aid feedback," J. Acoust. Soc. Am. **115**, 1147–1156 (2004).

[7]D. P. Egolf, B. T. Haley, K. A. Weaver, and D. S. Barker, "The hearing aid feedback path: Mathematical simulations and experimental verification," J. Acoust. Soc. Am. **78**, 1578–1587 (1985).

[8]D. P. Egolf, B. T. Haley, H. C. Howell, and S. Legowski, "Simulating the open-loop transfer function as a means for understanding acoustic feedback in hearing aids," J. Acoust. Soc. Am. **85**, 454–467 (1989).

[9]K. Nakao, R. Nishimura, and Y. Suzuki, "Calculation of transfer function of acoustic feedback path for in-the-ear hearing aids with a correction for specific acoustic impedance of a tubule," Acoust. Sci. & Tech. **27**, 242–244 (2006).

[10]J. Kates, "Room reverberation effects in hearing aid feedback cancellation," J. Acoust. Soc. Am. **109**, 367–378 (2001).

[11]J. Maxwell and P. Zurek, "Reducing acoustic feedback in hearing aids," IEEE Trans. Speech Audio Process. **3**, 304–313 (1995).

[12]J. Kates, "Feedback cancellation in hearing aids: A computer simulation results," IEEE Trans. Signal Process. **39**, 553–562 (1991).

[13]M. M. A. Spriet, I. Proudler, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," IEEE Trans. Signal Process. **53**, 3749–3763 (2005).

[14]J. Hellgren, "Analysis of feedback cancellation in hearing aids with filtered-x lms and the direct method of closed loop identification," IEEE Trans. Acoust. Speech Signal Process. **10**, 119–131 (2002).

[15]R. J. Oliveira, "The active earcanal," J. Am. Acad. Audiol. **8**, 401–410 (1997).

[16]G. Ma, F. Gran, F. Jacobsen, and F. Agerkvist, "A new approach for modelling the dynamic feedback path of digital hearing aids," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (2009), pp. 209–212.

[17]Y. Jingbo, T. Mengtong, and J. Chang, "Modeling external feedback path of an itedigital hearing instrument for acoustic feedback cancellation," in 2005 IEEE International Symposium on Circuits and Systems (2005), pp. 1326–1329.

[18]J. Ehrenberg, T. Ewart, and R. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," J. Acoust. Soc. Am. **63**, 1861–1865 (1978).

[19]R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via mode-wrelax," IEEE Trans. Aerosp. Electron. Syst. **35**, 294–307 (1999).

[20]J. Li and R. Wu, "An efficient algorithm for time delay estimation," IEEE Trans. Signal Process. **46**, 2231–2235 (1998).

[21]R. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, Englewood Cliffs, NJ, 1973).

[22]G. Forsythe, M. Malcolm, and C. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, NJ, 1976).

[23]M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," IEEE Trans. Acoust. Speech Signal Process. **36**, 477–489 (1988).

[24]Product Data and Specifications of Kemar Manikin Type 45BA, G.R.A.S. Sound & Vibration, Holte, Denmark (2007).

[25]HI-PRO Hardware Installation Guide, GN Otometrics A/S, Taastrup, Denmark (1993).

[26]L. L. Beranek, *Acoustics* (McGraw-Hill, New York, 1954).

The upper limit of temporal pitch for cochlear-implant listeners: Stimulus duration, conditioner pulses, and the number of electrodes stimulated

Robert P. Carlyon and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Colette M. McKay
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 8 May 2009; revised 15 October 2009; accepted 18 December 2009)

Three experiments studied discrimination of changes in the rate of electrical pulse trains by cochlear-implant (CI) users and investigated the effect of manipulations that would be expected to substantially affect the pattern of auditory nerve (AN) activity. Experiment 1 used single-electrode stimulation and tested discrimination at baseline rates between 100 and 500 pps. Performance was generally similar for stimulus durations of 200 and 800 ms, and, for the longer duration, for stimuli that were gated on abruptly or with 300-ms ramps. Experiment 2 used a similar procedure and found that no substantial benefit was obtained by the addition of background 5000-pps "conditioning" pulses. Experiment 3 used a pitch-ranking procedure and found that the range of rates over which pitch increased with increasing rate was not greater for multiple-electrode than for single-electrode stimulation. The results indicate that the limitation on pulse-rate discrimination by CI users, at high baseline rates, is not specific to a particular temporal pattern of the AN response.
 [DOI: 10.1121/1.3291981]

PACS number(s): 43.66.Ts, 43.66.Hg [RYL] Pages: 1469–1478

I. INTRODUCTION

An important finding in the cochlear-implant (CI) literature concerns the ability of CI users to process differences in the repetition rate of an electrical stimulus, such as a pulse train. For pulse rates lower than about 300 pps, listeners can detect rate differences of a few percent, and some listeners can use pulse-rate differences to recognize melodies and to estimate or even produce musical intervals (Pijl and Schwarz, 1995; McDermott and McKay, 1997; Moore and Carlyon, 2005). However, at higher pulse rates, these abilities usually decline dramatically; although marked individual differences exist, the vast majority of CI users do not reliably associate increases in pulse rate above about 400 pps with increases in pitch, and in many listeners this "upper limit" is considerably lower (Shannon, 1983; Townshend *et al.*, 1987; McKay *et al.*, 2000; Zeng, 2002; but see Kong *et al.*, 2009).

The upper limit of temporal pitch in CI users is of both clinical and scientific importance. Clinically, attempts to encode "temporal fine structure" in CIs (Nie *et al.*, 2005; Stickney *et al.*, 2005; Riss *et al.*, 2008) are likely to be limited by the range of temporal repetition rates that can be processed by CI users. Scientifically, several authors have pointed out the apparent paradox between the inability of CI users to encode high pulse rates and both the accurate encoding of such stimuli in the auditory nerve of deafened animals, and evidence that normal-hearing (NH) listeners use phase locking cues to estimate the frequency of pure tones up to at least 2000 Hz (Moore, 1973; Hartmann *et al.*, 1990; Shepherd and Javel, 1997; Micheyl *et al.*, 1998). This in turn has led to suggestions that accurate temporal processing may be subjected to additional requirements that are not met by CI stimulation; these have included (i) stochastic activity in the auditory nerve (AN) (Rubinstein *et al.* 1999), (ii) comparison of timing differences across different AN fibers produced by the basilar-membrane (BM) traveling wave (Shamma, 1985; Moore and Carlyon, 2005), (iii) a "match" between the temporal pattern of activity in a given nerve fiber and the place on the BM that it innervates (Moore, 1982; Oxenham *et al.*, 2004; Moore and Carlyon, 2005), and (iv) stimulation of auditory nerve fibers that innervate very apical regions of the cochlea (Middlebrooks and Snyder, 2009).

Despite its importance, the locus of the high-rate limitation remains largely unknown. Circumstantial evidence for a peripheral limitation, at the level of the AN, comes from measures of the electrically evoked compound action potential (ECAP). Wilson *et al.* (1997a) reported that although the ECAP to each pulse of a low-rate (e.g., 100 pps) pulse train was of roughly equal size, at higher pulse rates (e.g., 1000 pps) the ECAP to odd-numbered pulses was larger than that to even-numbered pulses. This "alternating-amplitude" pattern was attributed to a large number of neurons being activated by the first pulse, refractory when the second pulse was presented, recovered by the third, and so on. It could affect pitch perception at high rates by conveying two intervals to the brain, corresponding to the time between every pulse and to that between every second pulse [Fig. 1(a)]. For example, if the compound activity of the AN were processed by an array of "more central" neurons having different thresholds (Carlyon *et al.*, 2008b), those with higher thresholds would respond only to the odd-numbered pulses and therefore convey a pulse rate an octave lower than that actually presented. If the depth of the alternating-amplitude pattern increased with pulse rate, then more and more of these neurons would

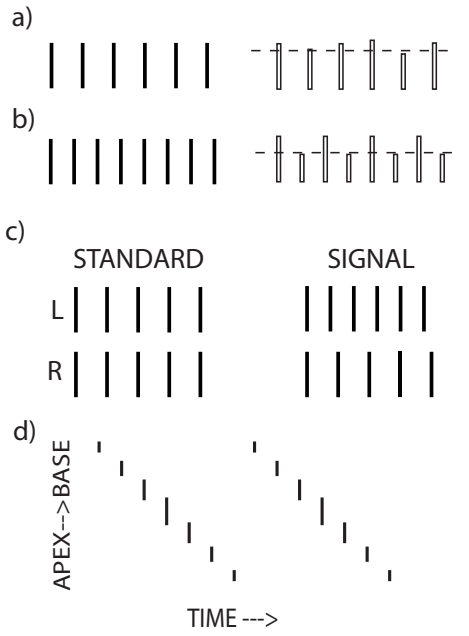


FIG. 1. (a) and (b) show schematic pulse trains (left-hand side, filled bars) together with hypothetical "alternating-amplitude" ECAPs (right-hand side, open bars). More central neurons having thresholds above the dashed lines would respond to every other pulse. If the ECAP becomes more modulated with increasing rate (b) then the increase in the number of more central neurons responding only to every other pulse might counteract the increase in pitch resulting from the shorter interpulse interval. (c) shows pulse trains presented in the standard and signal intervals of the "dichotic" conditions studied by van Hoesel and by Carlyon (van Hoesel and Clark, 1997; van Hoesel, 2007; Carlyon *et al.*, 2008a). The pulse rate presented to the left ear in the signal interval differs not only from that presented to the same ear in the standard interval but also from that presented to the right ear in the signal interval. The resulting binaural cue aids performance at low, but not at high, pulse rates (Carlyon *et al.*, 2008a). (d) represents the pulse trains presented on seven electrodes in the "peaked" condition of experiment 3.

convey rates corresponding to every second pulse, and this could counteract the increase in pitch expected from shorter interpulse intervals [Fig. 1(b)].

Evidence also exists for a more central limitation. Oxenham *et al.* (2004) required NH listeners to discriminate differences in the modulation rate of so-called "transposed tones" (Bernstein and Trahiotis, 2002)—high-frequency tones multiplied by a half-wave-rectified low-frequency sinusoid. They argued that these stimuli should produce similar patterns of AN activity to a pure tone, but in fibers having characteristic frequencies (CFs) corresponding to the carrier (but see Dreyer and Delgutte, 2006). Frequency difference limens (DLs) were considerably higher than for pure tones, and, importantly, listeners could not extract a "missing fundamental" from three harmonically related modulators each applied to a separate high-frequency carrier. They argued that the processing of phase locking at the level of the brainstem and above required either a match between the place and rate of stimulation or a comparison between the relative timing of activity in different AN fibers. More recently, Kong *et al.* (2009) showed that although the variation in rate discrimination as a function of baseline rate differed across listeners, for a given listener this variation was similar for rate discrimination of simple pulse trains and for the discrimination of the rate of sinusoidal amplitude modulation applied to a 5000-pps carrier. Presumably, the two types of stimulus produced different patterns of AN activity, tentatively suggesting that the across-listener differences may have arisen from more central processes.

The present study evaluates stimulus manipulations that would be expected to markedly affect the pattern of AN activity and which could feasibly extend the upper limit of temporal pitch. Experiment 1 examines the effect of increasing stimulus duration from 200 to 800 ms, and of turning the pulse trains on and off with long (300-ms) ramps. Physiological data indicate that the alternation in the ECAP amplitude to successive pulses gets smaller later in the pulse train (Wilson *et al.*, 1997b), so using a longer stimulus should provide a portion of the stimulus during which the alternation is reduced. Furthermore, we would expect the use of gradual onsets to also reduce this alternation because the effective "first pulse" will differ across neurons according to their threshold. Experiment 2 investigated whether the addition of a background 5000-pps pulse train would improve rate discrimination of moderate-rate (100–500 pps) pulses. The use of these so-called "conditioning pulses" has been shown to abolish the alternating-amplitude ECAP pattern (Rubinstein *et al.*, 1999). Experiment 3 studied whether pitch increased monotonically to higher overall pulse rates when multiple electrodes, rather than a single electrode, were stimulated (McKay *et al.*, 2005). Overall, none of the manipulations performed in the three experiments produced a substantial improvement in performance at the upper end of the rates studied. Section V discusses the generality of the present findings and considers additional steps that may be necessary to extend the range of rates that can be accurately encoded by CI users.

II. EXPERIMENT 1: EFFECTS OF SIGNAL DURATION AND ONSET/OFFSET ENVELOPE

A. Method

Six users of the Nucleus CI24M implant (Table I) were presented with biphasic pulse trains to electrode 11, in monopolar (MP1+2) mode, with a pulse duration of 45 μs per phase and an interphase gap of 8 μs. In different conditions, the pulse trains were either turned on abruptly and had a duration of 200 or 800 ms or, for the 800-ms duration, were turned on and off with 300-ms ramps. The ramps were linear in terms of Cochlear Corporation's (roughly logarithmic) clinical "current units" (CUs), with end points corresponding to threshold and the steady-state amplitude of the central portion of the pulse train (see below). Stimuli were controlled using the APEX software package (Laneau *et al.*, 2005) and checked using a test implant and digital storage oscilloscope.

Rate discrimination was measured for 100-, 200-, 300-, 400-, and 500-pps standards using the "mixed-block" procedure described by Kong *et al.* (2009). On each trial, the listener was presented with the standard and a signal, having a 35% higher rate, in random order, and was required to indicate which interval contained the higher pitch. Responses were scored as correct when the signal interval was chosen, and correct-answer feedback was provided. The order in

TABLE I. Details of the cochlear-implant users who took part in the experiments.

Subject	Age (years)	Etiology	Duration of profound deafness (years)	Duration of implant use	BKBq (%)	BKBn (%)
S1	61	CSOM	10	8 years	90	85
S2	77	Otosclerosis/noise induced	22	5 years	90	69
S3	41	Congenital progressive	35	2 years	97	46
S4	76	Progressive unknown	5	5 years	84	78
S5	66	Sudden, unknown	21	6 years	89	56
S6	71	Otosclerosis	>30	7 years	49	22
S7	60	Sudden, viral	>30	1 year	58	14
M1	44	Progressive unknown	2 years	18 months	98.7%	88.0%
M2	72	Unknown	Unknown	2 years	Unknown	Unknown

which the different standard rates were presented from trial to trial was randomized.[1] Each block consisted of 10 trials per standard rate, and the results from 10 blocks were averaged; each data point presented here therefore corresponds to the mean of 100 trials.

Prior to the main part of the experiment, the stimuli were loudness balanced using the method described by Landsberger and McKay (2005). A pair of stimuli was presented to each subject. At the initial presentation, the first stimulus in the pair was fixed at "comfort ("C") level" and the second was presented at a much quieter level. The subject was instructed to raise the level of the variable sound (second stimulus) until it was slightly louder than the fixed sound and then reduce the level until the two sounds were equally loud. All baseline rates were loudness matched to the baseline rate of 100 pps, and then each signal was loudness matched to the corresponding baseline.

B. Results

Data for each listener are shown separately in the six panels of Fig. 2(a). Listeners S1–S5 show the "classic" pattern in most conditions, with very good performance at the lowest standard rates, that deteriorates to near-chance levels at the highest rates tested. Exceptions are listener S1, who performs close to ceiling at all rates tested for the 200-ms duration, and listener S3, who for the 200-ms duration shows the nonmonotonic pattern recently described by Kong *et al.* (2009). Listener S6 shows an unusual pattern, in all three conditions, with below-chance performance at intermediate rates. Note that this occurred despite the fact that correct-answer feedback was provided after every trial. This result is consistent with a decrease in pitch with increasing pulse rate over some range of rates, and we will return to this finding in Sec. IV.

Data averaged across listeners are plotted in Fig. 2(b), which shows that mean performance declines monotonically with increasing standard rate. Although performance appears to be slightly better at high rates for the 200-ms duration, a two-way (rate × condition) repeated-measures analysis of variance (ANOVA) found no interaction between rate and condition [$F(8,5)=1.6$, $p=0.2$], and the main effect of condition just failed to reach significance [$F(2,10)=3.7$, $p=0.06$].

The rationale for experiment 1 was that the alternating-ECAP pattern should be reduced by turning the pulse trains on gradually and that the reduction in the alternation observed later in the pulse train in animal experiments might reduce the influence of the alternation on perception, by allowing the listener to focus on the later portions of the stimulus. However, neither of these two manipulations helped. Overall, performance at high rates was, if anything, slightly better for the shorter (200-ms) stimulus, and the addition of 300-ms ramps produced only a small and nonsignificant improvement (compare upright and inverted triangles in Fig. 2).

III. EFFECT OF HIGH-RATE "CONDITIONING" PULSES

A. Rationale

As noted in the Introduction, it has been suggested that the deterministic nature of the AN response to electrical stimulation may impair the encoding of temporal information. Two approaches—the addition of noise to the signal and the addition of high-rate background pulses—have been proposed (Morse and Evans, 1996; Rubinstein *et al.*, 1999; Zeng *et al.*, 2000; Chatterjee and Robert, 2001). In both cases, it has been suggested that there is an optimal level of the added stimulus that increases the stochasticity of the response to the signal without "swamping" it. Experiment 2 was motivated by suggestion of Rubinstein *et al.* (1999) that a more stochastic pattern of firing could be introduced by the addition of high-rate background pulses, and by their demonstration that the alternating-amplitude ECAP pattern in response to a 1016-pps pulse train could be eliminated by the addition of a 5081-pps conditioner. Note that, although the high-rate pulse trains used to study the effects of conditioners on threshold and on neural responses are typically presented continuously (Runge-Samuelson *et al.*, 2004; Hong and Rubinstein, 2006), this is not crucial for an effect on the ECAP amplitude alternation to be observed; the conditioner used by Rubinstein *et al.* (1999) was turned on only 29 ms before the 1016-pps pulse train. Experiment 2 used the same "mixed-block" procedure as in experiment 1, with the following differences.

Four users of the Nucleus CI24M implant and two users of the MedEl Pulsar implant (M1 and M2, see Table I) took part. In both cases the stimulus duration was 300 ms. For the CI24M users, rate discrimination was measured for baseline

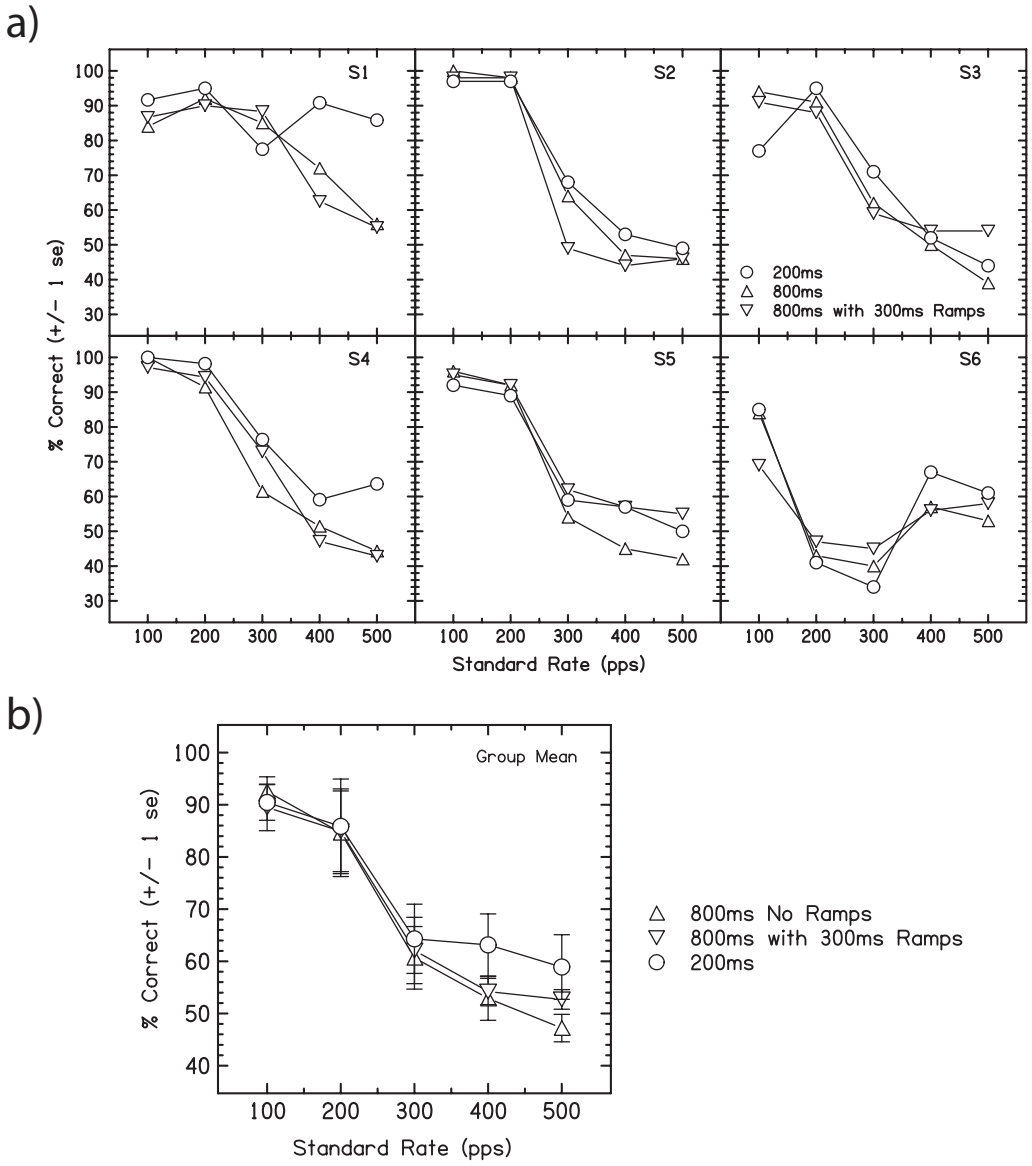


FIG. 2. (a) shows rate discrimination, obtained using the "mixed rate" procedure, for the six subjects of experiment 2. Data are shown for signal durations of 200 and 800 ms, and, in the latter case, with and without 300-ms onset and offset ramps. (b) shows the data averaged across listeners, with standard errors indicated by error bars.

rates of 100–500 pps in the absence of a conditioner and in the presence of a 5000-pps conditioner that, in different conditions, could have one of a range of different levels. These levels were based on a preliminary experiment that measured the threshold and comfort ("C") level for a 300-ms 5000-pps pulse train. The conditioner levels were then chosen to be at −20%, 0%, 20%, 40%, 60%, and 80% of the dynamic range, where 0% corresponds to threshold and 100% to C level. These percentages were calculated in terms of Cochlear Corporations CUs, which are a roughly logarithmic function of current. The conditioners were turned on 300 ms before and off 200 ms after, each standard and signal pulse train. Interpulse intervals were filled with an integer number of conditioner pulses at the conditioner rate. The different conditioner levels were run in separate blocks of 10 trials per point, in a counterbalanced order, until 100 trials per point had been collected for each subject and condition. Pulse duration was 45 μs/phase and the interphase duration was 8 μs. Stimuli were checked as in experiment 1.

Stimulation of the MedEl implant was implemented using the "RIB2" software and hardware produced by the University of Innsbruck, and checked by connecting the output of a "detector box" whose internal electronics are the same as for a Pulsar implant, to a digital storage oscilloscope. The experimental method was similar to that for the CI24M users, except that dynamic range was defined in terms of MedEl's (approximately linear) current units, and that the conditioner pulse trains were turned on 4 min before each block of trials.[2] For these listeners, within each session two blocks of ten trials/point were run for each conditioner level before moving to the next-highest level, to avoid any residual effects from the conditioner in one block of trials on the neural responses during the next block. In addition, one listener (M1) was subsequently retested about 1 month later using conditioners turned on and off before each stimulus, as for the CI24M users. Pulse duration was 43 μs/phase with no interphase gap. For both groups of listener, the stimuli were loudness balanced in the absence of a conditioner using

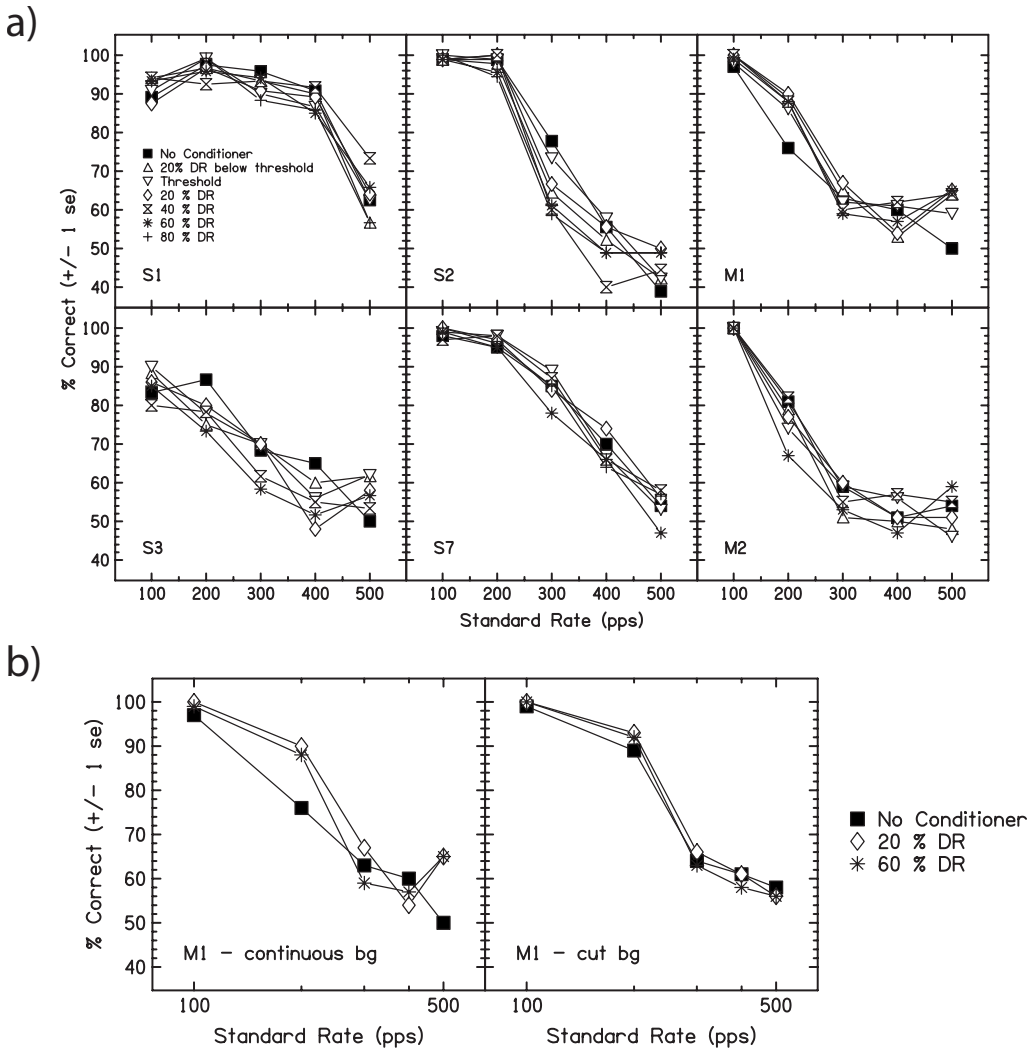


FIG. 3. (a) shows rate discrimination data from experiment 2, for conditions without a conditioner (solid squares) and for conditioners of various levels (open symbols and asterisks; color online). (b) shows the original data for listener M2 obtained with a continuous conditioner in the left-hand panel, together with, in the right-hand panel, a retest in which the conditioner was turned on and off before each stimulus.

the same method as for experiment 1. These loudness balanced levels were used regardless of the presence or level of the conditioner.

B. Results

Results from the six listeners who took part in experiment 2 are shown in Fig. 3(a). Performance in the absence of a conditioner, shown by the filled squares, was universally good at the lowest rate tested (100 pps) and declined monotonically at higher standard rates. Generally speaking, the conditioner pulses had no substantial effect on performance. Minor exceptions were listeners S2 and S3, for whom most conditioners impaired performance, and listener M1, for whom all levels of conditioner produced slightly better performance at 200 and 500 pps than when no conditioner was present. To test whether this small improvement depended on the conditioner pulses being presented continuously, this listener was retested several months later with conditioners that were turned on 300 ms before and off 200 ms after, each standard and signal pulse train. The retest was performed with no conditioner and with conditioner levels corresponding to 20% and 60% of the dynamic range. The results are shown in the right-hand panel of Fig. 3(b), with the original data for the same conditions replotted in the left-hand panel. The main difference between the two plots is that performance at 200 and 500 pps, in the absence of the conditioner, was better in the second set of measures. Combined with the fact that performance for all levels of the continuous conditioner was similar, the most parsimonious explanation for the apparent effect of the continuous conditioner is that the initial estimate of no-conditioner performance was, for some reason, too low at 200 and 500 pps, and that there was no reliable effect of adding a conditioner.[3]

IV. EXPERIMENT 3: CONCURRENT STIMULATION OF MULTIPLE ELECTRODES

A. Rationale and method

McKay *et al.* (2005) investigated an informal clinical observation that, when a patient's CI is reprogramed to use a speech-processing strategy with a faster pulse rate, the patient sometimes reports hearing a change in pitch even though both the original and new pulse rates are much higher

than 300 pps. They considered the possibility that the reported pitch change arose from the change in current level necessitated by the new rate, which may have resulted in "place-of-excitation" differences. To test this idea, they generated nine seven-electrode stimuli, comprising three "spectral profiles" (with current maxima at apical, middle, and basal locations) and three pulse rates of 450, 900, and 1800 pps per electrode. A multidimensional scaling study revealed separate perceptual dimensions for spectral profile and pulse rate. The authors concluded that the increases in pulse rate above 450 pps, when applied to multiple electrodes, produced detectable changes in a perceptual dimension that were not conveyed by place-of-excitation cues. This dimension could, of course, be pitch, but other possibilities exist, including the greater adaptation observed for higher rates of stimulation (Zhang *et al.*, 2007) and changes in other temporal properties such as perceived roughness.

Experiment 3 tested whether increases in pulse rate produce increases in pitch over a wider range with multiple electrodes, compared to single-electrode, stimulation. In the single-electrode condition, 500-ms 25 μs/phase pulse trains applied to electrode 11 of a CI24M implant in monopolar ("MP1+2") mode. The interphase gap was 25 μs. Nine pulse rates, ranging from 112.5 to 1800 pps in half-octave steps, were used. In the "flat" and "peak" spectral profile conditions, the same pulse rates were applied concurrently to electrodes 8–14 inclusive. Electrodes were stimulated in a basal to apical order and the interpulse interval between electrodes corresponded to 1/8 of the stimulus period. Before testing, comfort (C) levels were obtained for each subject, spectral profile, and rate. Initially, subjects indicated the comfort level of each electrode presented in isolation and at a rate of 450 pps. For multiple-electrode configurations, C level was determined by presenting all seven electrodes at their respective relative levels and at a rate of 450 pps, initially at a level well below expected comfort level. For "flat" spectral profiles, each electrode was presented at the C level determined for single-electrode presentation minus a global adjustment (in CUs) compensating for the combined loudness effects of several electrodes. "Peaked" profiles were based on these same individual electrode levels but with attenuations of 25, 18, 5, 0, 5, 18, and 25 CUs applied to electrodes 8–14, respectively. The levels of all electrodes were then varied together until the subject reported that the C level had been reached. Using the C level obtained at 450 pps as a reference, the C level for 112.5 225, 900, and 1900 pps trains were obtained by the method of adjustment described above.

For each condition, the pitches of the different pulse-rate stimuli were compared in experiment 3a using the "Midpoint comparison" (MPC) procedure described by Long *et al.* (2005). The MPC procedure ranks stimuli along a single perceptual dimension (e.g., pitch) in an optimally efficient manner, by means of paired comparisons. Initially, two stimuli are selected at random and the subject indicates which has the higher pitch. Next, a new stimulus is compared to the higher ranked of these two and, if judged higher, is put in first place; otherwise it is then compared to the lower-ranked stimulus and positioned either second or third. As an example, later on in the procedure, a provisional ranking of seven stimuli might be [C I E A F B D]. The new stimulus, G, would first compared to the middle rank (A), and if judged higher the list is bisected so that it would be compared to B; if judged lower than B, it would then be compared to F. The procedure is continued until a rank is obtained for all electrodes and was repeated 20 times for each condition, allowing us to estimate a mean and standard error for each rank. For each repetition, the various rates were introduced to the procedure in a fresh random order. The procedure has several desirable properties, that it shares with the mixed-block procedure used in experiments 1 and 2: (i) it is not subjected to many of the biases that plague magnitude-estimation ("pitch scaling") tasks (Poulton, 1979); (ii) there is no single stimulus that is presented repeatedly throughout each block, as is the case with adaptive procedures or the method of constant stimuli; this reduces the likelihood of the subject focusing on some idiosyncratic feature of one stimulus; (iii) pitch varies substantially throughout each run, helping the subject focus on that dimension; and (iv) each run includes some easily discriminable pitches, preventing the subject from becoming discouraged. We chose to use it instead of the mixed-block procedure for three reasons: (i) the absence of feedback makes the procedure more appropriate for revealing pitch reversals, (ii) it provides an overview of how pitch varies over a wide range of closely spaced rates, and (iii) because we were studying a wide range of rates, many of which we expected to produce very similar pitches, the mixed-block procedure, in which only adjacent rates are compared, would result in the subject spending much of the time guessing. This problem was partially alleviated by the midpoint comparison procedure because nonadjacent rates were often compared.

To determine whether the different pulse rates were discriminable, experiment 3b used the mixed-block procedure described in Sec. II, with the important difference that a three-interval two-alternative forced-choice (3I2AFC) trial structure was used. This "odd-man-out" procedure allowed the subject to use any difference between the stimuli to identify the stimulus with the higher rate, which could occur in either the second or third intervals. Both single-electrode and peaked stimuli were used, with standard rates of 112.5, 225, 450, 900, and 1272 pps. Signal rates were always half an octave higher than the corresponding standard, and correct-answer feedback was provided after every trial.

B. Results

The results of the MPC procedure of experiment 3a are shown in Fig. 4(a). For the single-electrode condition (filled triangles), the results are generally consistent with those obtained using the mixed-block procedure of experiments 1 and 2. Listeners S1, S2, S4, and S7 all show an increase in pitch up to some value, above which the functions asymptote. Furthermore, listener S2, whose functions asymptote at a lower rate than the others, also showed discrimination performance that declined at a low rate in experiments 1 and 2. Listeners S1 and S7, whose pitch rankings increased up to rates of about 636 and 318 pps, respectively, also showed good per-

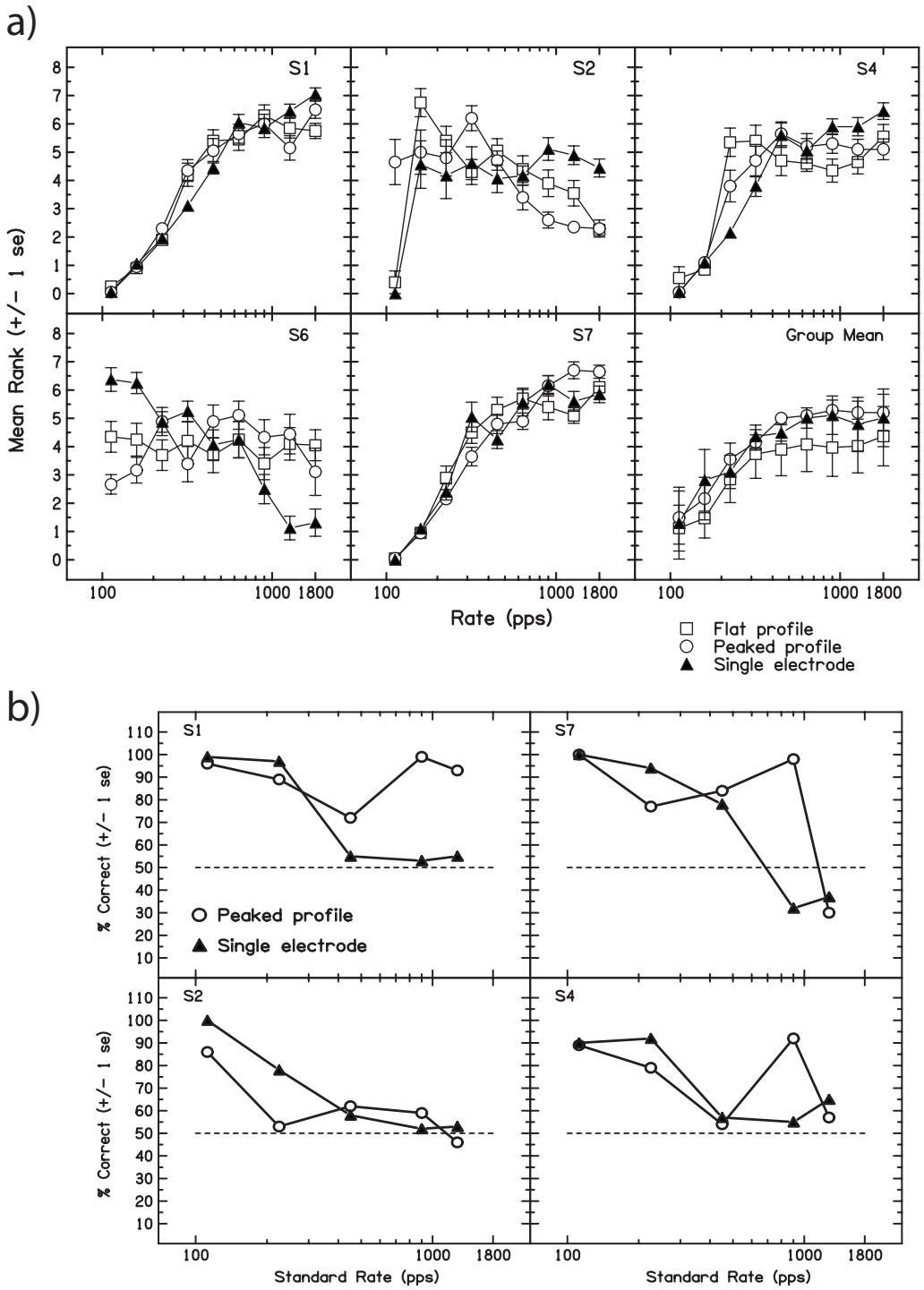


FIG. 4. (a) shows the results of the midpoint-comparison procedure of experiment 3a for single-electrode and multielectrode stimulation. (b) shows the results of the odd-man-out-discrimination task used in experiment 3b.

formance up to relatively high rates in experiment 2 [Fig. 3(a)]. Listener S6 shows an unusual pattern, in that pitch appears to decrease with increasing pulse rate between 159 and 225 pps, and also above 636 pps. This listener also showed an unusual pattern in experiment 1 [Fig. 2(a)], with below-chance performance for standard rates of 200 and 300 pps.

The topic of primary interest concerns the comparison between the functions obtained with single-electrode and multielectrode stimuli. Inspection of Fig. 4(a) shows that in no case does pitch continue to increase up to a higher rate for the multielectrode (open symbols) stimuli than for the single-electrode (filled triangles) stimulus. The only discernable effects are a lower asymptote for listener S4, flat functions for S6, and a decrease in pitch at higher pulse rates for S2. A two-way repeated-measures ANOVA on the mean ranks for each subject, excluding the anomalous data of S6, revealed the expected main effect of pulse rate [$F(8,24)=11.78$; $p<0.02$], but no main effect of spectral profile [$F(2,6)=1.0$, n.s.] and no interaction [$F(16,48)=1.22$, n.s.].[4]

The discrimination results from the odd-man-out task of experiment 3b are shown in Fig. 4(b). For three listeners (S1, S7, and S4), there are one or more high rates where discrimination is better with multielectrode (peaked) than with single-electrode stimulation, even though pitch ranking over these range of rates had reached an asymptote in experiment 3a. This may help reconcile the negative findings of experiment 3a with observation of McKay *et al.* (2005) that changes to pulse rate applied to multiple electrodes gave rise to a discriminable perceptual dimension, even though rate changes to a single electrode were not detectable in a forced-choice task. That is, adding electrodes may allow a rate change to become discriminable, without necessarily providing a consistent change in pitch. One way in which this could happen comes from finding of Zhang *et al.* (2007) that adaptation in AN fibers depends on pulse rate: with interleaved stimulation on different electrodes the "effective" pulse rate for any given neuron is increased, and this could result in more adaptation. Alternatively, multielectrode stimulation could influence the temporal pattern of firing, e.g., by abolishing the alternating-amplitude ECAP pattern, and this difference could be detectable without improving pitch perception. Another possible explanation is that subjects used small loudness differences in the forced-choice procedure, which did not lead to differences in pitch, and that the salience of these differences depended on the number of electrodes being stimulated.

V. DISCUSSION

A. Neural basis of temporal pitch perception by CI users

The results presented here show that a number of stimulus manipulations, that one would reasonably expect to have a substantial impact on the AN response to a pulse train, do not produce a marked improvement in rate discrimination at high rates. A related finding was obtained by Kong *et al.* (2009), who compared rate discrimination of regular pulse trains with the ability to discriminate changes in modulation rate imposed on a 5000-pps carrier. They found that, in both conditions, the variation in performance as a function of the standard (modulation) rate differed markedly across listeners, but that, for a given listener and electrode, the pattern of results was similar for the two stimuli. Similarly, in experiment 2 of the present study, the upper limit differed across listeners in a way that was immune to the addition of conditioning pulses of a range of levels. Overall, the results lead to the conclusion that the upper limit of temporal pitch is not specific to any particular temporal pattern of AN activity. However, this does not, of course, mean that features of the auditory nerve response can never affect pitch perception. Indeed, we have argued elsewhere that, at low rates, neural refractoriness does affect temporal pitch (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003; Carlyon *et al.*, 2008b), and the presence of some pitch reversals in experiment 3 could result from an alternating-amplitude ECAP without this particular pattern being the reason why temporal pitch generally asymptotes at high rates.

Although the present data argue against an explanation of the upper limit in terms of peripheral phenomena such as reflected by the alternating-amplitude ECAP pattern, data from other studies indicate that the limitations to high-rate processing are not specific to pitch judgments. For bilateral CI users, discrimination of interaural time differences between pulse trains presented to each ear also deteriorates with increases in pulse rate (van Hoesel and Clark, 1997; van Hoesel and Tyler, 2003; van Hoesel, 2007), and, even when a rate difference between matched electrodes in the two ears provides a usable binaural cue at low rates, this is not the case when the pulse rate is high enough for monaural rate discrimination to break down [Carlyon *et al.*, 2008a: see Fig. 1(c) of the present article]. Hence we can tentatively conclude that the upper limitation arises at a site of processing that is central to the AN but common to binaural and pitch processes. One possibility is suggested by tantalizing observation of Snyder *et al.* (1995) that cells in the inferior colliculus (IC) differ in the highest rate to which they will phase lock, but that this "upper limit" is similar, for a given cell, regardless of whether the phase locking is to the pulse rate of a simple pulse train or to the modulation rate imposed on a high-rate carrier. Hence a stimulus-independent limitation may occur at or before the IC, although we should of course stress that this suggestion remains speculative.

The data presented here and elsewhere are also consistent with the idea, mentioned in the Introduction, that the electrical stimulus would need to be modified in order for more central processes to effectively process the temporal activity present in the AN. As long ago as 1982, Moore (1982) suggested that the range of interpulse intervals that is effectively processed by the brain is linked to the characteristic frequencies of the AN fibers that convey that information, and this idea has been more recently implemented in a quantitative model (Bernstein and Oxenham, 2005). Interestingly, although this idea suggests that high-rate information should be more effectively conveyed by stimulating the base of the cochlea, recent recordings from the IC suggest that, conversely, stimulation of AN fibers that innervate the apex of the cochlea may better convey temporal information (Middlebrooks and Snyder, 2009). Another suggestion is that the brain may exploit the phase differences in the response of AN fibers having different CFs, which arise from the slowing of the traveling wave on the BM (Shamma, 1985; Moore and Carlyon, 2005; Cedolin and Delgutte, 2007). Clearly, the stimuli used here and in other studies have neither reproduced these timing differences nor produced either an exact "place-rate" match or stimulation of apical AN fibers. Hence the present results should not be interpreted as meaning that *no* stimulus manipulation can improve temporal processing of high-rate electrical stimuli. It is also true that none of our listeners fell into the "star" category who can reliably discriminate rates higher than 500 pps (Hochmair-Desoyer *et al.*, 1983; Wilson *et al.*, 1997b; Kong *et al.*, 2009). It is therefore possible that either our method of stimulation or the neural status of our particular listeners imposed some additional limitation on performance that masked any potential improvement in the peripheral neural code. This issue is currently under active investigation in our laboratory (Macherey and Carlyon, 2009). What is clear, however, is that for

typical CI users, and in the absence of special types of stimulation that are yet to be implemented, manipulations that we would expect to substantially change the temporal pattern of auditory nerve activity do not have a marked or consistent effect on temporal pitch perception at high rates.

ACKNOWLEDGMENT

We thank Antje Ihlefeld for programming the RIB2 system used in experiment 3.

[1]As discussed by Kong *et al.* (2009), this procedure has several strengths: (i) unlike an adaptive procedure, it is possible to measure performance in conditions where it is likely to be close to chance, (ii) the variation in the standard from trial to trial helps keep the listener's attention focused on the pitch dimension and encourages him/her to use that perceptual dimension for all judgments, (iii) the presence of some "easy" (low-rate) standards in each block prevents discouragement, and (iv) the use of a two-interval, rather than odd-man-out procedure, makes it less likely that the listener will use some extraneous cue such as loudness or roughness differences. A weakness is that performance at some rates may be obscured by ceiling effects. However, this was considered not to be critical for the present study, whose purpose was to study performance at rates that, in some conditions, were expected to be close to chance.

[2]For listener M1, all conditioners having a level above threshold (0% of dynamic range) were still audible at the end of this 4-min period, and no marked reduction in loudness was observed. Listener M2 described conditioner levels of 20% and 40% as just audible at the start and inaudible at the end of the 4-min adaptation period, with some reduction in loudness for the 60% level

[3]The conclusion that conditioner pulses did not improve rate discrimination was supported by the results of a two-way repeated-measures ANOVA. The use of this statistic is complicated by the facts that the "optimal" conditioner listener may vary across listeners (Hong and Rubinstein, 2006) and that different conditioner levels were used for the Cochlear and MedEl listeners. We therefore used only two levels of the "conditioner" variable: no conditioner at all, and the level that, for each listener, gave the highest percent-correct score averaged across all rates. Such a method runs the risk of obtaining a significant effect of conditioner by chance because we are selecting the conditioner level producing the best performance from a set of six. Nevertheless, the ANOVA, while producing a main effect of rate $[F(4,20)=15.575;\ p<0.001]$, produced no effect of conditioner $[F(1,5)=2.96,$ n.s.] and no significant interaction $[F(4,20)=1.86,$ n.s.].

[4]When all subjects were included, there was still no effect of profile and no interaction, but the main effect of rate just failed to reach significance.

Bernstein, J. G. W., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

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

Carlyon, R. P., Long, C. J., and Deeks, J. M. (**2008**a). "Pulse-rate discrimination by cochlear-implant and normal-hearing listeners with and without binaural cues," J. Acoust. Soc. Am. **123**, 2276–2286.

Carlyon, R. P., Mahendran, S., Deeks, J. M., Long, C. J., Axon, P., Baguley, D., Bleeck, S., and Winter, I. M. (**2008**b). "Behavioral and physiological correlates of temporal pitch perception in electric and acoustic hearing," J. Acoust. Soc. Am. **123**, 973–985.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Cedolin, L., and Delgutte, B. (**2007**). "Spatio-temporal representation of the pitch of complex tones in the auditory nerve," 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. 61–70.

Chatterjee, M., and Robert, M. E. (**2001**). "Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?" J. Assoc. Res. Otolaryngol. **2**, 159–171.

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.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–1724.

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.

Hong, R. S., and Rubinstein, J. T. (**2006**). "Conditioning pulse trains in cochlear implants: Effects on loudness growth," Otol. Neurotol. **27**, 50–56.

Kong, Y.-Y., Deeks, J. M., Axon, P. R., and Carlyon, R. P. (**2009**). "Limits of temporal pitch in cochlear implants," J. Acoust. Soc. Am. **125**, 1649–1657.

Landsberger, D. M., and McKay, C. M. (**2005**). "Perceptual difference between low and high rates of stimulation on single electrodes for cochlear implantees," J. Acoust. Soc. Am. **117**, 319–327.

Laneau, J., Boets, B., Moonen, M., van Wieringen, A., and Wouters, J. (**2005**). "A flexible auditory research platform using acoustic or electric stimuli for adults and young children," J. Neurosci. Methods **142**, 131–136.

Long, C. J., Nimmo-Smith, I., Baguley, D. M., O'Driscoll, M., Ramsden, R., Otto, S. R., Axon, P. R., and Carlyon, R. P. (**2005**). "Optimizing the clinical fit of auditory brain stem implants," Ear Hear. **26**, 251–262.

Macherey, O., and Carlyon, R. P. (**2009**). "Effect of intracochlear stimulation site on the upper limit of temporal pitch," in Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.

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., Henshall, K. R., and Hull, A. E. (**2005**). "The effect of rate of stimulation on perception of spectral shape by cochlear implantees," J. Acoust. Soc. Am. **118**, 386–392.

McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). "Place and temporal cues in pitch perception: Are they truly independent?," ARLO **1**, 25–30.

Micheyl, C., Moore, B. C. J., and Carlyon, R. P. (**1998**). "The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones," J. Acoust. Soc. Am. **104**, 1039–1050.

Middlebrooks, J., and Snyder, R. (**2009**). "Enhanced transmission of temporal fine structure using penetrating auditory nerve electrodes," in Association for Research in Otolaryngology, 32nd Midwinter Research Meeting, Baltimore, MD.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Springer Handbook of Auditory Research: Pitch Perception*, edited by C. J. Plack and A. J. Oxenham (Springer, New York), pp. 234–277.

Morse, R. P., and Evans, E. F. (**1996**). "Enhancement of vowel coding for cochlear implants by addition of noise," Nat. Med. **2**, 928–932.

Nie, K. B., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Pijl, S., and Schwarz, D. W. F. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.

Poulton, E. C. (**1979**). "Models for the biases in judging sensory magnitude," Psychol. Bull. **86**, 777–803.

Riss, D., Arnoldner, C., Baumgartner, W. D., Kaider, A., and Hamzavi, J. S. (**2008**). "A new fine structure speech coding strategy: Speech perception at a reduced number of channels," Otol. Neurotol. **29**, 784–788.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). "Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation," Hear. Res. **127**, 108–118.

Runge-Samuelson, C. L., Abbas, P. J., Rubinstein, J. T., Miller, C. A., and Robinson, B. K. (**2004**). "Response of the auditory nerve to sinusoidal electrical stimulation: Effects of high-rate pulse trains," Hear. Res. **194**,

1–13.
Shamma, S. (**1985**). "Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.
Shepherd, R. K., and Javel, E. (**1997**). "Electric stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status," Hear. Res. **108**, 112–144.
Snyder, R., Leake, P. A., Rebscher, S. J., and Beitel, R. (**1995**). "Temporal respolution of neurons in cat inferior colliculus to intracochlear electrical stimulation: Effects of neonatal deafening and chronic stimulation," J. Neurophysiol. **73**, 449–466.
Stickney, G. S., Nie, K. B., and Zeng, F. G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.
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 Hoesel, R. J. M. (**2007**). "Sensitivity to timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.
van Hoesel, R. J. M., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.
van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. (**2003**). "Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing," J. Acoust. Soc. Am. **114**, 1516–1528.
Wilson, B., Finley, C., Lawson, D., and Zerbi, M. (**1997**a). "Temporal representations with cochlear implants," Am. J. Otol. **18**, s30–s34.
Wilson, B., Zerbi, M., Finley, C., Lawson, D., and Honert, C. V. D. (**1997**b). "Speech processors for auditory prostheses (Eighth Quarterly Progress Report)," NIH Project No. N01-DC-5-2103.
Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zeng, F. G., Fu, Q. J., and Morse, R. (**2000**). "Human hearing enhanced by noise," Brain Res. **869**, 251–255.
Zhang, F., Miller, C. A., Robinson, B. K., Abbas, P. J., and Hu, N. (**2007**). "Changes across time in spike rate and spike amplitude of auditory nerve fibers stimulated by electric pulse trains," J. Assoc. Res. Otolaryngol. **8**, 356–372.

Binaural masking level differences in actual and simulated bilateral cochlear implant listeners

Thomas Lu[a)]
Department of Otolaryngology–Head and Neck Surgery, University of California, Irvine, California 92697

Ruth Litovsky
Department of Communicative Disorders, Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin 53705

Fan-Gang Zeng
Department of Otolaryngology–Head and Neck Surgery, University of California, Irvine, California 92697

(Received 11 May 2009; revised 13 December 2009; accepted 15 December 2009)

At present commercially available bilateral cochlear implants (CIs) improve their users' speech understanding in noise but they employ two independent speech processors that cannot provide accurate and appropriate interaural level and time differences as seen binaurally in normal hearing (NH) listeners. Previous work suggests that binaural cues are accessible to bilateral CI users when presented to single pairs of pitch-matched electrodes, but the scope was limited and the mechanisms remained unclear. In this study, binaural masking level differences (BMLDs) were measured in five bilateral Nucleus-24 CI users over multiple pairs of pitch-matched electrodes. Average BMLD was 4.6 ± 4.9 dB, but large individual variability prevented significance ($p=0.09$). Considering just the 125 Hz condition, as in previous work, phase (N0S0 vs N0Sπ) and electrode effects were significant. Compared with simulated bilateral CI users, actual bilateral CI users had proportionally higher thresholds for N0Sπ than N0S0. Together the present results suggest that the performance gap in BMLDs between CI and NH listeners is not due to a lack of sufficient acoustic cues in the temporal envelope domain but to a true binaural deficit related to a central mechanism in deprived binaural processing. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290994]

PACS number(s): 43.66.Ts, 43.66.Pn, 43.64.Me, 43.66.Qp [BLM] Pages: 1479–1490

I. INTRODUCTION

Cochlear implant (CI) users can understand most speech in quiet, but the presence of noise can greatly challenge that understanding (i.e., Zeng and Galvin, 1999; Spahr and Dorman, 2004; Stickney *et al.*, 2004). Although a second implant can benefit CI users for hearing in noise (Müller *et al.*, 2002; Litovsky *et al.*, 2004, 2006a; Buss *et al.*, 2008; Litovsky *et al.*, 2009), binaural sensitivity in these bilateral CI users falls short of normal hearing (NH) levels (van Hoesel *et al.*, 1993; Buss *et al.*, 2008; Nopp *et al.*, 2004; Laszig *et al.*, 2004; van Hoesel, 2004; Verschuur *et al.*, 2005; Neuman *et al.*, 2007; van Hoesel *et al.*, 2008). In addition, there is a performance gap between testing the binaural sensitivity of CI users under ideal circumstances (i.e., Long *et al.*, 2006) and their performance under real world conditions (i.e., van Hoesel *et al.*, 2008). So far, measures of true binaural hearing, such as squelch, which requires computation of differences between sounds entering the left and right ears, appear to be of marginal benefit to bilateral CI users (Nopp *et al.*, 2004; Laszig *et al.*, 2004; Long *et al.*, 2006; van Hoesel, 2004; Buss *et al.*, 2008; van Hoesel *et al.*, 2008). Uncoordinated stimulation timing produced by independent speech processors may be one limiting factor for binaural sensitivity. However, data from experiments using the Spear3 research interface (Hearworks, Pty, Melbourne, Australia), which can deliver highly synchronized stimulation pulses to both ears, still show lower than normal levels of binaural function in CI (Long *et al.*, 2006; van Hoesel *et al.*, 2008; Litovsky *et al.*, 2010).

Binaural masking level difference (BMLD) is one measure of binaural sensitivity (Hirsh, 1948). Unlike head shadow cues or redundancy, BMLD requires precise computations of differences in the sounds entering the left and right ears (Durlach, 1963). For a signal-in-noise detection task, when the signal phase is inverted between the two ears, there is a higher likelihood for detection than when the signal phase is identical between the two ears. BMLD is calculated as the difference in threshold between the two conditions. In NH listeners this difference can be as large as 20 dB for tones (van de Par and Kohlrausch, 1997). Under similar stimuli conditions, BMLDs measured in bilateral CI users averaged ~9 dB when stimulating with single pairs of electrodes (Long *et al.*, 2006). Similar tests in children show 6.4 dB for CI and about 10 dB for NH (Van Deun *et al.*, 2009a, 2009b). The cause of this gap between normal and CI listeners is not well understood.

Furthermore, BMLD in CI users appears to be much smaller for more complex sounds. An experiment that used speech under realistic listening conditions reported <1.5 dB of binaural unmasking (van Hoesel *et al.*, 2008), which is significantly less than the 9 dB found in Long *et al.*, 2006. In contrast, NH listeners typically had BMLD values that fell

[a)]Author to whom correspondence should be addressed. Electronic mail: telu@uci.edu

between the range seen for tones and speech (Levitt and Rabiner, 1967). This would suggest that in NH listeners binaural processing of BMLD for speech can be predicted by BMLD for tones. In the present study, we examined several factors that could potentially affect BMLDs in bilateral CI users under more realistic stimulation conditions than previously studied.

The first factor is that BMLD values may be influenced by electrode position along the cochlea. Complex stimuli such as speech have multiple frequency components, and in normal hearing listeners, BMLD decreases with increasing signal frequency (van de Par and Kohlrausch, 1997). BMLDs were shown to be similar between a 125 Hz sinusoid and a 125 Hz envelope in the form of "transposed" stimuli with a 4000 Hz carrier. This "transposition" of the stimuli places the 125 Hz signal at the 4000 Hz cochlear region. The similarity in BMLDs for NH listeners suggests that the 125 Hz sinusoid and the 125 Hz envelope of the transposed stimuli were processed similarly despite the activity at different cochlear regions (125 Hz vs 4 kHz). This is relevant to continuous interleaved sampling (CIS) based processing strategies in CI as sounds are typically transmitted via envelopes on top of a high pulse rate carrier (Wilson *et al.*, 1991). For CI users, BMLDs were reported for basal, middle, and apical electrode pairs, but only one pair was tested in each subject (Long *et al.*, 2006). Although BMLDs are expected to be similar over multiple cochlear regions in a single CI user as they are in NH listeners, it has not yet been experimentally verified.

A second factor that may influence BMLD levels in CI is the stimulus signal frequency. In normal hearing listeners, ~10 dB BMLDs were reported with signal frequencies as high as 4000 Hz (van de Par and Kohlrausch, 1997). Existing studies show that CI users have difficulty discriminating higher pulse rates and modulation frequencies starting around 300 Hz under monaural (Zeng, 2002) and binaural (Carlyon *et al.*, 2008) conditions. The 125 Hz signal frequency used in Long *et al.*, 2006 is well within the temporal discrimination ability of CI users, but BMLD at higher signal frequencies has not been systematically addressed. If the mechanisms that cause this limitation play a role in binaural sensitivity, BMLD values may deteriorate with higher stimulus frequencies and with complex sounds that have higher frequency components.

Third, there is the possibility that the interaction between narrowband noise and a 125 Hz signal combined with the envelope extraction by the speech processor produces enough detectable difference cues between NmSm and NmS-m, where m and -m indicate monaural and the phase reversed monaural noise and signal stimuli. If these NmSm and NmS-m are discriminable monaurally, it could inflate BMLD values. For this reason, it is important to test whether BMLDs could be a result of monaural discrimination of the signal envelope.

Fourth, prior auditory deprivation can negatively impact binaural processing (Silman *et al.*, 1984; Hall and Grose, 1993; Gray *et al.*, 2009). It is known that the lack of auditory inputs can lead to neural degeneration in the periphery (Hinojosa and Lindsay, 1980). Furthermore, there is evidence, at least in children, that binaural processing continues to develop with experience and age (Hogan *et al.*, 1996; Litovsky, 1997; Hall *et al.*, 2004; Litovsky *et al.*, 2006b), and a loss of binaural input could arrest this development.

A recent study of patients with unilateral hearing loss due to atresia suggested that there is a loss of 2 dB binaural advantage due to squelch for each decade of life with only unilateral hearing (Gray *et al.*, 2009). The recovery of binaural sensitivity was not studied. However, it was earlier reported that adult patients with unilateral conductive hearing loss experience a blunting of binaural sensitivity, which eventually recovers to normal or near-normal levels by 1 year after surgical intervention (Hall and Grose, 1993). These subjects had normal cochlear function and once the conductive hearing loss was remedied, their hearing function was essentially restored to normal. Despite restoring the ability to hear, the cochlear implants do not restore normal hearing due to coarse frequency resolution, and furthermore, the use of independent speech processors virtually guarantees uncoordinated stimulation in pulse timing, thus providing abnormal interaural time difference (ITD) cues in the fine-structure of sound. Therefore even if the auditory system is capable of recovering normal binaural function, it would not be able to do so given abnormal and coarse frequency inputs provided by both CIs.

Auditory deprivation on unilateral cochlear implant use has been documented (Tyler and Summerfield, 1996), the effects of which on binaural function in adults has not been adequately explored and there are limited studies. The main effect is that some binaural function such as squelch is recovered over the course of 1 year (Buss *et al.*, 2008). The evidence for bilateral hearing aids is a little clearer. Comparing two groups, one fitted unilaterally with hearing aids and the other bilaterally, there was a difference between ears in the unilaterally implanted group (Silman *et al.*, 1984). As a result of auditory deprivation, it is unknown to what extent limited binaural processing is due to the implant and speech processor and how much is due to degeneration of binaural processing in the central auditory system. This may be addressed, in part, by comparing the performance of NH listeners under acoustic CI simulation to that of actual CI users.

The overall aim of this study is to investigate the extent of binaural sensitivity quantifiable by BMLD in bilateral CI users and why it falls short of normal performance on binaural listening tasks. Building upon the results from Long *et al.* (2006), the questions that this study will directly address are as follows: (1) How BMLDs compare across basal, middle, and apical pairs of pitch-matched electrodes; (2) how BMLDs change with signal frequency; (3) would masking level differences under monaural conditions predict binaural BMLDs; and (4) how do normal hearing listeners compare when subjected to acoustic CI simulations.

II. METHODS

A. Subjects

Five bilaterally implanted Nucleus 24 patients were tested (Table I). Several of these patients wore hearing aids for a number of years. The two implantations were spaced 1

TABLE I. CI subjects.

	Gender	Age	Age at deafness	Etiology	Age at hearing aid use	Left/right implanted (years ago)	Basal pair (L/R)	Middle pair (L/R)	Apical pair (L/R)	HINT speech scores quiet/noise (%)
CI1	f	62	26	Ototoxicity	26	6/18	5/5	12/12	20/21	L:83/32 R:95/65 Bi: NA
CI2	m	51	3	Congenital	3	2/3	5/5	12/12	22/22	L:0/0 R:0/0 Bi:0/not tested
CI3	f	63	37	Meniere's	43	4/5	5/5	12/10	20/20	L:23/not tested R:91/93 Bi 99/96
CI4	f	72	26	Otosclerosis	38	3/4	6/6	11/16	14/16	L: 91/29 R:95/11 Bi:99/40
CI5	m	68	34	Progressive	42	6/6	5/5	13/12	20/20	L: 98/39 R: 98/55 Bi: 100/58

year apart in three subjects and 12 years in one subject. The fifth subject had both ears implanted simultaneously. Speech performance scores for these subjects are listed in Table I. Sound input levels were set so that maximum stimulation levels were approximately 95% of most comfortable level (MCL) for that electrode pair.

For each subject, pairs of pitch-matched electrodes were found and balanced for loudness at MCL since binaural sensitivity is higher for pitch-matched pairs in CI (van Hoesel and Clark, 1997; Lawson *et al.*, 1998; Long *et al.*, 2003, 2007) and frequency-matched stimuli in NH (Henning, 1974; Nuetzel and Hafter, 1981). As an example, an apical pair (i.e., 20L, 20R) was selected and threshold and comfort levels were determined individually. Each electrode in a pair was played sequentially and the subject was asked to make a pitch judgment. If there was a perceived pitch difference, the next adjacent electrode on one side was selected, typically the more recently implanted ear, and the pitch was compared again. This search for a pitch-matched pair was iterated as many times as necessary until a suitable match was found. If needed, the loudness was balanced by lowering the MCL of the louder electrode to match. The pitch and loudness match was confirmed again. In a few cases, subjects reported that the similarity of pitch match changed when balancing loudness, requiring a search for a different pitch-matched pair. Three pairs of electrodes—apical, middle, and basal—were tested in each subject. The exact locations of these matched pairs are listed in Table I. Finding and balancing each pair of pitch-matching electrodes could take anywhere from 30 min to over 1 h, and sometimes longer.

B. Stimulus and hardware configuration

All stimuli were generated digitally using MATLAB (Mathworks, Inc., Natick, MA) with 44.1 kHz sampling rate through a PC sound card. Stimuli consisted of 300 ms sinusoids at 125, 250, or 500 Hz. Masking sounds were 400 ms band-passed noise with bandwidths of ±25, 50, and 100 Hz, centered at each sinusoid, respectively. N0S0 signals denote masking noise that was identical for both ears, and sinusoids that had a zero-phase difference (also identical) between the two ears. N0Sπ denotes the same masking noise, but sinusoids that are 180° out of phase (inverted) between the two ears. The ratio of amplitude of the sinusoid to the noise is the signal to noise ratio (SNR) in decibel.

The signal processing algorithm is similar to that described in Long *et al.*, 2006. The stimuli were presented through line-in jack, bypassing the microphone, on the Spear3 research interface (Hearworks, Pty, Melbourne, Australia) running a custom program that implemented a bilateral one-channel CIS program. Audio signals were digitized by the on-board analog-to-digital converters. The signal was half-wave rectified and low-passed filtered using a four-pole Butterworth filter. For the 125 and 250 Hz stimulus conditions, the signal was low-passed at 500 Hz, and the envelope of the signal was used to modulate a pulse train with a rate of 1000 pulses per second (pps). For the 500 Hz stimulus condition, the signal was low-passed at 1000 Hz before modulating a 2000-pps pulse train. A standard loudness growth map was used to compress the amplitude of the signal (as in Long *et al.*, 2006). The stimulation mode used was MP1+2 (monopolar), with 25 μs per phase and an 8 μs pulse gap. The left and right pulsatile outputs of the Spear3 were verified to be synchronized to within 1 μs by viewing signals on an oscilloscope after being fed through an "implant-in-a-box" (Cochlear Pty, Lane Cove, Australia).

C. Implementation of acoustic CI simulation

Ten normal hearing subjects (20–26 years old) were tested using unprocessed stimuli along with several different

vocoder configurations to acoustically simulate a bilateral CI listening condition. Only two subjects (NH1 and NH2) who were tested under unprocessed conditions returned for testing with the vocoder. A single-channel noise-excited vocoder was used to determine the extent to which envelope cues were preserved. Inputs were identical to those used for the experiments with CI listeners. Stimuli were half-wave rectified and low-passed at 500 Hz to extract the envelope. The resulting signal was used to modulate a broadband noise. Since the one-channel noise vocoder used here does not accurately simulate the conditions in which basal, middle, or apical pairs of electrodes were stimulated, a second vocoder model based on Gaussian-enveloped tones (Lu *et al.*, 2007) was used to simulate synchronized and pulsatile stimulation for different electrode pairs. As with the noise vocoder, stimuli were half-wave rectified and low-pass filtered. The resulting envelope signal was used to modulate the amplitude of tone pulses that had Gaussian-shaped time-amplitude envelopes. The pulse rate was set at 1000 Hz for the Gaussian-enveloped tone pulses. The duration of each pulse was approximately 1 ms. The simulated basal, middle, and apical channels had center frequencies of 5367, 1426, and 369.5 Hz, based on Greenwood's map (Greenwood, 1990).

Gaussian-enveloped tones have been used in binaural psychophysical experiments (Buell and Hafter, 1988; van den Brink and Houtgast, 1990), but its application in a vocoder has not been reported previously. It is similar to the transposed tones used in van de Par and Kohlrausch, 1997. Two important characteristics of cochlear implants, which can be modeled with the vocoder, are not accounted for in current noise-based CI simulations. The first is pulsatile stimulation. With CI the signal envelope is sampled at the pulse rate. Noise-excited vocoders have a continuous representation of the signal envelope, which at low stimulation rates, the deficiency to simulate well separated pulses becomes readily apparent. The spread of excitation can be controlled by the duration of the pulse, which can widen or narrow the spectral spread produced by each pulse. Since ITD sensitivity is better at lower stimulation rates, and pitch perception is affected by the spread of excitation, the ability to control these factors makes the Gaussian-enveloped tone vocoder a potentially useful platform for acoustically simulating bilateral CI use.

D. Testing procedure

Masking level thresholds were estimated by a three-interval, forced-choice adaptive procedure with a two-down/one-up decision rule (Levitt, 1971). The custom software was implemented in MATLAB (Mathworks, Inc., Natick, MA). All three intervals contained N0. The signal S0 or Sπ was added to a single, randomly assigned interval; thus the target interval was either N0S0 or N0Sπ. The subject's task was to identify which one of the three intervals contained the tone (N0S0 or N0Sπ vs N0). Two correct successive responses decreased the SNR for the next trial, while one incorrect response immediately increased the SNR for the next trial. Each change in direction of SNR (decreasing to increasing or vice versa) counted as a reversal. After eight total reversals, with stepsize decreasing from 5 to 1 dB after the initial three reversals, the SNR at which the last four reversals occurred were averaged to produce an estimate of the SNR threshold. Smaller values indicate better detection of the signal in noise compared to a larger SNR. Each run of the test consisted of two, randomly interleaved tracks, N0S0 and N0Sπ to help control for subject state between the two conditions. Once eight total reversals were observed in one track, the program presented stimuli exclusively from the other track until eight total reversals were also observed. With every subject, three threshold estimates were obtained for each phase (N0S0 vs N0Sπ), signal frequency, and electrode pair combination.

Data for monaural conditions were obtained by disconnecting either the left or right transmitting coil from the subject's head and repeating the same test protocol described above. Normal hearing subjects were tested in a sound-proof chamber using calibrated headphones (HDA 200, Sennheiser, Old Lyme, CT). Sound levels were set to 70 dB sound pressure level (SPL), calibrated to a 1 kHz tone having the same rms level as the stimuli used in the experiment.

E. Data analysis

Average values were calculated for N0S0 and N0Sπ. The BMLD was calculated as the overall difference between the average N0S0 and N0Sπ values. An unpaired t-test was used to determine whether the three thresholds from N0S0 were statistically different from the three N0Sπ thresholds, with $p<0.05$ being the criterion for significance (see asterisk symbols).

For the pooled data, average thresholds and BMLD were calculated over all stimulus frequencies, electrode pairs, and subjects. A repeated-measures analysis of variance (r.m. ANOVA) was performed on the thresholds to determine the effect of phase (N0S0 and N0Sπ), electrode (basal, middle, and apical), and signal frequency (125, 250, and 500 Hz). A second r.m. ANOVA was repeated on just the data set containing 125 Hz.

III. RESULTS

A. Binaural masking level differences

Masking level thresholds for various combinations of signal frequency and electrodes for N0S0 and N0Sπ are shown in Fig. 1. Masking level thresholds over subjects and conditions were all above 0 dB for N0S0, ranging from 0.3 to 11.6 dB SNR (signal to noise ratio), and both below and above 0 dB SNR (-10.6 to 12.5 dB) for N0Sπ. In fact, thresholds in the N0Sπ condition only reached negative values, where the amplitude of the signal was less than that of the noise, in two subjects (CI1 and CI5). On an individual basis, masking thresholds were generally lower for N0Sπ compared to N0S0, indicating a BMLD (see Table II).

BMLD values are shown in Fig. 2, where vertical bars show the BMLD calculated for each electrode pair grouped by stimulus. Asterisks indicate statistically significant BMLDs (unpaired t-test, $p<0.05$), that is, higher average thresholds for N0S0 than N0Sπ. This comparison helps to identify cases in which there existed a meaningful difference

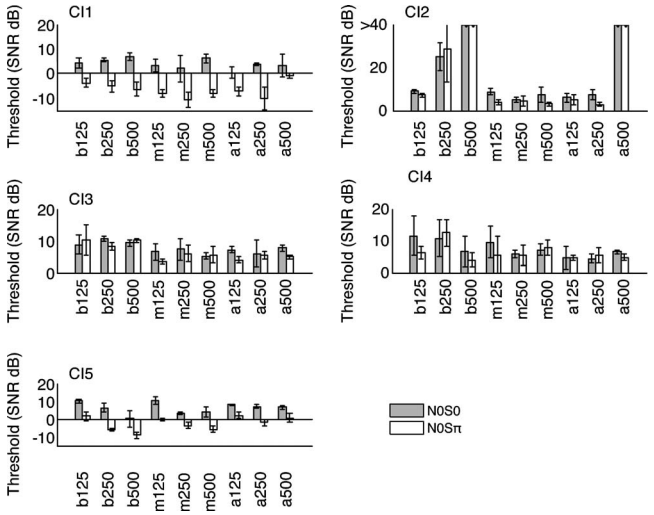


FIG. 1. Masking level thresholds for CI. Masking level thresholds are indicated by each bar. These indicate the smallest signal-to-noise ratio (in decibel) at which the subject was able to detect the signal in noise (see Sec. II). Each group represents a different electrode pair and signal frequency condition. The labels designate the electrode pair (base/middle/apex) and stimulus condition (125/250/500 Hz). For example, a250 indicates apical pair of electrodes and stimulus condition of 250 Hz. In each group, the shaded bars represent the average of three trials from N0S0 while the unshaded bars designate the N0Sπ condition. The error bars represent standard deviation. For subject 2, some data points were truncated at 40 dB as the subject was unable to complete the signal detection task under those conditions.

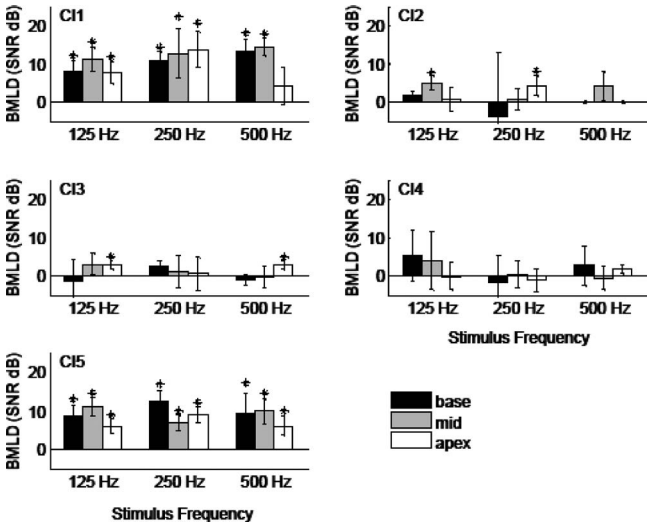


FIG. 2. Binaural masking level differences from five CI subjects. Each bar represents a BMLD value (y-axis) averaged over three trials, and is the mean difference of the N0S0 and N0Sπ thresholds in Fig. 1. BMLDs are grouped by stimulus condition along the x-axis. Data from different electrode pairs are indicated by the shading. The error bars represent the standard deviation. "$*$" denote statistically significant BMLD values (unpaired t-test, $p<0.05$).

in performance between the two conditions, regardless of the magnitude of the difference (i.e., small but significant vs large and insignificant BMLDs). For CI1 and CI5, BMLDs were statistically significant over almost all basal, middle, and apical electrode pairs and stimulus conditions. Average BMLDs (mean $\pm$ s.d.) were 10.7 ± 3.4 and 8.8 ± 2.2 dB, respectively, for CI1 and CI5. BMLDs observed for CI2 and CI3 were generally lower (mean$=1.5 \pm 2.8$ and 1.2 ± 1.8 dB, respectively), with only two out of nine conditions showing statistically significant BMLDs. CI4 showed no statistically significant BMLDs (mean$=1.2 \pm 2.4$ dB) under any stimulus conditions.

Over all subjects and conditions tested, the average BMLD was 4.6 ± 4.9 dB. The main effect of phase on thresholds (N0S0 vs N0Sπ) was not significant (ANOVA, $F_{1,4}=4.954$, $p=0.09$). There was no significant effect detected of either electrode or rate (ANOVA, $p>0.05$). The lack of significance in this case is most likely due to the large range of performance across individuals. It is worth noting that for each subject, thresholds for N0S0 were always higher than N0Sπ, but the standard deviations were always the same or higher for N0Sπ (see Table II).

When only the 125 Hz condition was considered, as in Long *et al.*, 2006, four of five subjects in the present study showed a significant BMLD in at least one cochlear region or electrode pair, with the average being 4.9 ± 3.8 dB. Results from repeated-measures ANOVA showed a significant

TABLE II. Average thresholds and BMLDs over all stimulus conditions and electrode pairs.

Subject	N0S0 (dB SNR)	N0Sπ (dB SNR)	BMLD (dB SNR)	R NmSm R NmS-m	L NmSm L NmS-m
CI1	3.9 ± 2.1	-6.7 ± 3.1	10.7 ± 3.4	3.1 ± 2.2 3.6 ± 2.3	3.2 ± 3.0 1.7 ± 1.1
CI2	16.4 ± 14.6	15.0 ± 16.3	1.5 ± 2.8	28.3 ± 14.5 25.0 ± 16.6	25.7 ± 15.9 26.3 ± 15.6
CI3	7.6 ± 1.7	6.4 ± 2.5	1.2 ± 1.8	22.7 ± 14.8 24.8 ± 12.9	23.38 ± 18.2 23.39 ± 18.2
CI4	7.4 ± 2.6	6.2 ± 2.6	1.2 ± 2.4	24.2 ± 17.4 24.3 ± 17.3	34.3 ± 6.6 31.2 ± 10.3
CI5	6.2 ± 3.3	-2.6 ± 3.9	8.8 ± 2.2	6.8 ± 1.6 6.6 ± 2.1	5.8 ± 2.6 6.5 ± 1.8

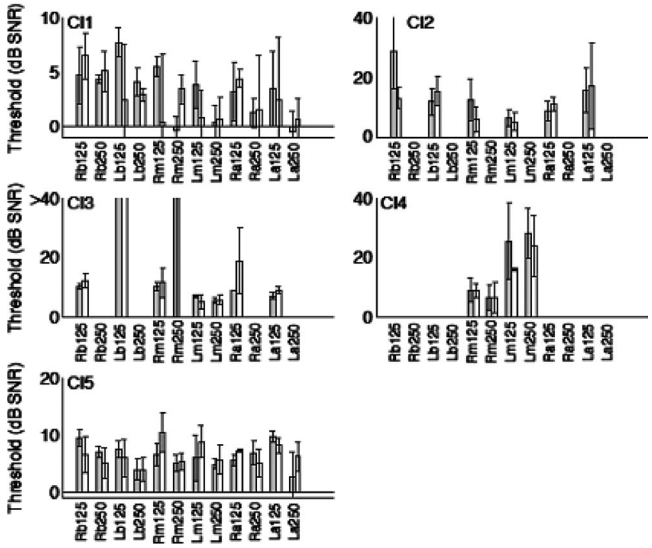


FIG. 3. Monoaural masking level thresholds. Format is similar to Fig. 1. The labels along the *x*-axis include "R" and "L" to specify right or left ear for the monaural condition. Note that some bars representing high threshold values were truncated (e.g., CI3).

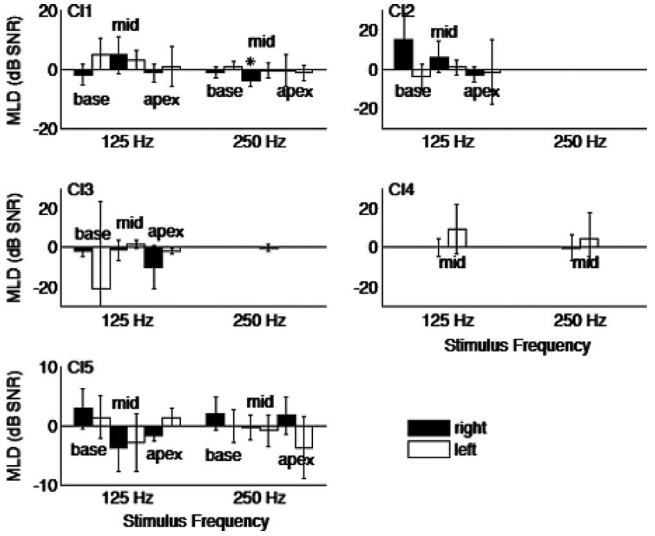


FIG. 4. Monaural MLDs. Format is similar to Fig. 2. For each group, shaded bars indicate data from the right ear, while unshaded bars are data from the left ear. Within each group, the left two bars are data from basal pair of electrodes, the third and fourth bars are from the middle pair of electrodes, and the last two bars are from the apical pair of electrodes.

effect of phase ($F_{1,4}=10.130$, $p=0.033$) and electrode ($F_{2,8}=13.366$, $p=0.003$), but no interaction between the two factors ($F_{2,8}=3.844$, $p=0.068$). *Post-hoc* t-tests with Bonferroni correction showed a significant difference ($p=0.023$) between average N0S0 (mean$=7.1\pm3.2$ dB) and N0Sπ (mean$=2.24\pm5.2$ dB) thresholds. Pair-wise comparisons of the electrodes revealed a significant difference ($p=0.023$) between thresholds of apical (mean$=3.41\pm4.4$ dB) and basal (mean$=6.50\pm4.7$ dB) electrodes. There were no significant differences between the thresholds of the middle electrode (4.2 ± 5.5 dB) and either apical ($p=0.82$) or basal ($p=0.71$) electrodes. Although electrodes had an effect on threshold, there was no significant effect on BMLD as a function of cochlear region (r.m. ANOVA, $F_{2,8}=3.842$, $p=0.068$).

Separate analysis of the 250 (mean$=4.6\pm5.9$ dB) and 500 Hz (mean$=4.5\pm5.1$ dB) conditions did not yield any statistical significance (ANOVA, $p>0.05$). Despite the large individual variability, these data suggest that at least some CI subjects can take advantage of binaural processing over multiple electrode pairs and signal frequency conditions.

B. Monaural detection thresholds

This study was also concerned with whether subjects can detect differences in amplitude modulation using a single ear. It is known that CI users can have very high sensitivity to amplitude modulations (Shannon, 1992). The performance under monaural conditions can reveal the contribution of binaural processing compared to what can be detected monaurally. Figure 3 shows masking level difference (MLD) thresholds from each ear separately for all five subjects. Only CI1 and CI5 were able to complete all test conditions, and thresholds for subjects C2–C4 were capped at 40 dB. Absolute thresholds tended to be higher than with the binaural condition, and MLDs between the two stimulus conditions, NmSm vs NmS-m, were not much different, where Nm indicates monaural noise, Sm is the monaural sinusoid, and S-m is the inverted monaural sinusoid. Under monaural conditions, no significant MLDs (unpaired t-test, $p>0.05$) were observed in any of the five subjects except for a single condition (CI1, 250 Hz) (Fig. 4).

Table II lists the average values obtained for monaural conditions. For CI1 and CI5, thresholds obtained for both right and left ears under monaural conditions are similar to the data from N0S0 condition, but not N0Sπ. These data from the monaural tests demonstrate that the BMLDs observed for the binaural condition cannot be explained by parallel but independent signal detection (via amplitude modulation of the signal envelope) between the left and right ears.

C. Binaural unmasking with acoustic simulation of CI

Because a prolonged lack of auditory input can have a negative impact on the normal functioning of the auditory system, the present manipulations were also conducted in normal hearing listeners using acoustic stimuli that were either unprocessed or with CI simulations. This approach would enable a distinction of issues stemming from CI-specific limitations related to the BMLD phenomenon. Thresholds, measured using unprocessed sounds, were on average 1.6 ± 3.5 dB for N0S0 and -6.3 ± 5.8 dB for N0Sπ (Fig. 5). BMLDs in unpracticed listeners ranged from 1.5 ± 1.1 to 13.9 ± 1.3 dB SNR (Fig. 6), which had a slightly wider range than that for CI listeners. Average BMLD was 7.9 ± 4.7 dB (see Table III for individual values). There was a significant effect of both phase (ANOVA, $F_{1,5}=19.855$, $p=0.007$) and signal frequency ($F_{2,10}=35.494$, $p\ll0.05$), but no significant interaction. Considering that these NH subjects had minimal practice, the range data reported here are still consistent with that reported in other studies (i.e., van de Par and Kohlrausch, 1997).

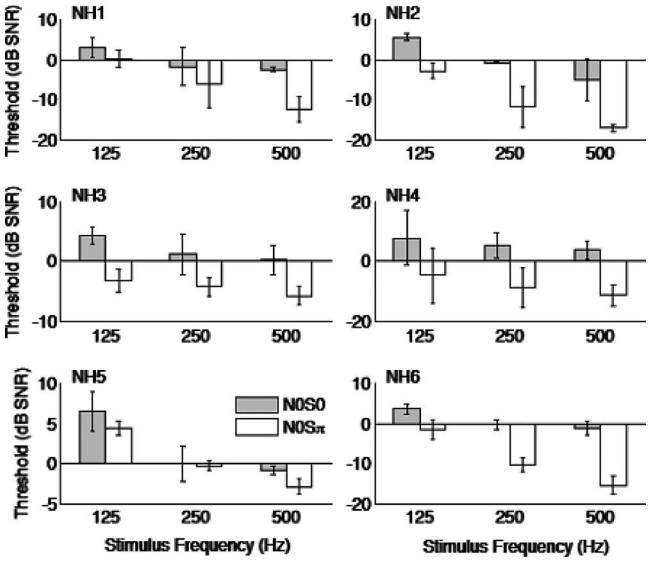


FIG. 5. Masking thresholds for unprocessed stimuli in NH listeners. Format is the same as Fig. 1.

When tested with a noise-excited, one-channel acoustic CI simulation, thresholds tended to be higher than in the unprocessed condition and quite variable (Fig. 7). Average thresholds were 12.2 ± 6.7 dB for N0S0 and -2.2 ± 12.9 dB for N0Sπ. This indicated that subjects tended to have much more difficulty with N0S0. However, subjects showed a large drop in thresholds with N0Sπ as seen in NH7 and NH8 for the noise-excited vocoder with 125 Hz signal condition. Average BMLD was 14.4 ± 14.7 dB. The effect of phase approached but did not reach statistical significance (ANOVA, $F_{1,5}=6.538$, $p=0.051$), and signal frequency had no significant effect ($F_{1,5}=1.179$, $p=0.327$).

Since using a single-channel noise vocoder does not accurately simulate the single-electrode stimulus conditions with electrically pulsed signals under which the CI users were tested, a vocoder based on Gaussian-enveloped tones was used to test simulated effects of basal, middle, and apical

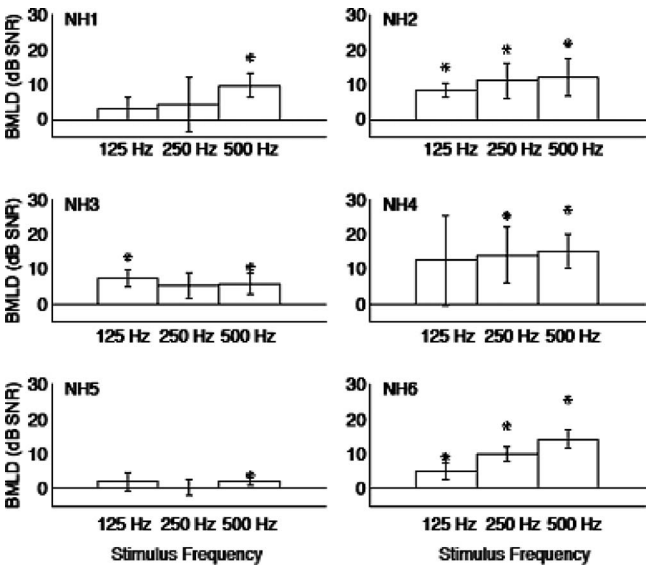


FIG. 6. BMLD for unprocessed stimuli in NH listeners. Format is the same as Fig. 2.

TABLE III. Average BMLDs over all stimulus conditions and electrode pairs for NH listeners.

Subject	Unprocessed (dB SNR)	Noise vocoder (dB SNR)	Gaussian vocoder (dB SNR)
NH1	5.8 ± 3.6	8.5 ± 1.2	8.9 ± 3.1
NH2	10.6 ± 1.9	9.3 ± 4.5	11.7 ± 6.3
NH3	6.2 ± 1.1	⋯	⋯
NH4	13.9 ± 1.3	⋯	⋯
NH5	1.5 ± 1.1	⋯	⋯
NH6	9.8 ± 4.6	⋯	⋯
NH7	⋯	26.6 ± 4.0	19.3 ± 4.0
NH8	⋯	21.8 ± 15.4	15.9 ± 9.8
NH9	⋯	28.0 ± 2.7	16.4 ± 4.2
NH10	⋯	-7.9 ± 14.4	3.1 ± 5.8

pair stimulations (Fig. 8). Thresholds for the Gaussian-enveloped tone vocoder averaged 4.0 ± 4.1 dB for N0S0 and -8.6 ± 7.2 dB for N0Sπ. Phase showed a statistically significant effect (ANOVA, $F_{1,5}=27.080$, $p=0.003$) but there was no effect of channel ($F_{2,10}=2.898$, $p=0.102$) or signal frequency ($F_{1,5}=2.772$, $p=0.157$). In addition, there were no significant interactions. Average BMLD was 12.6 ± 7.8 dB. Although BMLD values from the Gaussian-enveloped tone vocoders were similar to the noise vocoder condition, average N0S0 threshold was nearly 8 dB higher for noise vocoder.

A comparison of the N0S0 and N0Sπ threshold values from CI and NH using vocoder stimuli is shown in Fig. 9. Each data point represents a single-electrode pair and signal frequency combination from the five CI subjects and six NH subjects. The mean values for CI subjects were 6.6 ± 2.8 dB for N0S0 and 1.6 ± 6.1 for N0Sπ. The mean values for simulated CI (Gaussian vocoders) with NH listeners were 4.0 ± 4.1 dB for N0S0 and -8.6 ± 7.2 for N0Sπ. The mean values for unprocessed stimuli with NH listeners were 1.6 ± 3.5 dB for N0S0 and -6.3 ± 5.8 for N0Sπ. These mean values of N0S0 were significantly different from each other (unpaired t-test, $p < 0.05$). Mean values of N0Sπ were also significantly different (unpaired t-test, $p \ll 0.05$), except between the simulated CI and unprocessed stimuli in NH listeners (unpaired t-test, $p=0.26$). Despite the large individual variability, Fig. 9 illustrates three important points. First, CI implant users tended to produce the highest (or worst) thresholds in both binaural conditions, particularly in the N0Sπ condition. Second, the CI simulations captured the best performance seen in the CI user group, but not the poorest performance (three of five subjects in the present study). Third, the similar performance between CI simulations and unprocessed conditions in NH listeners suggests that the acoustic cues in the envelope are adequate to support normal BMLD performance. These results have important implications for CI signal and neural processing and will be discussed next.

IV. DISCUSSION

The purpose of this study was to establish BMLDs using an expanded set of electrodes and signal frequencies in order help understand why BMLDs are limited in bilateral CI users

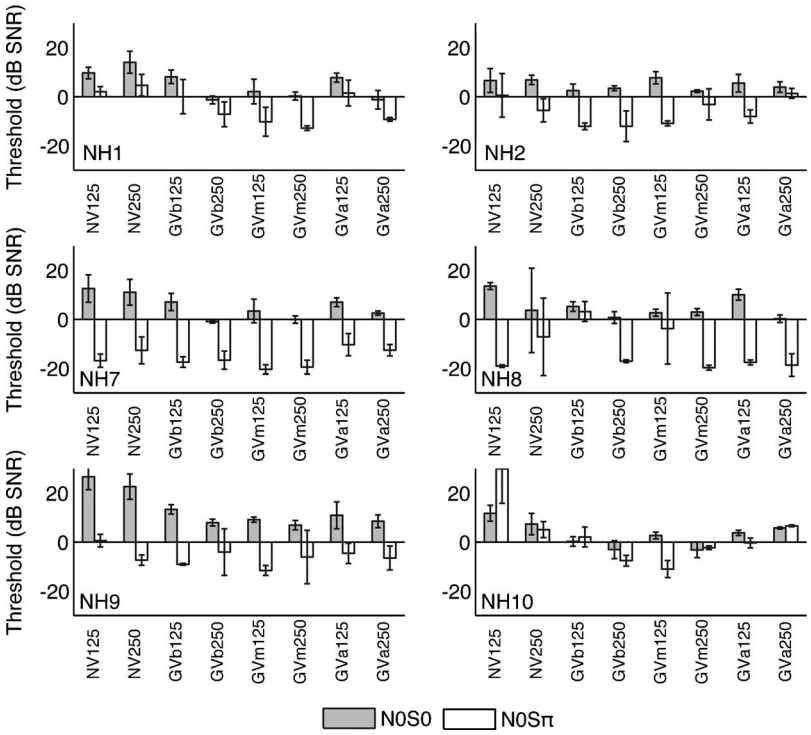


FIG. 7. Masking thresholds for NH listeners using acoustic CI simulation. Format is the same as Fig. 1. Labels along the *x*-axis include NV-noise-excited vocoder, GVb—Gaussian-enveloped tone vocoder (basal), GVm (middle), and GVa (apical). See Sec. II for details on the Gaussian-enveloped tone vocoder.

compared to NH listeners. The main findings of this study are that (1) BMLDs were measurable at multiple electrode locations for various signal/masker conditions, (2) BMLDs were more likely to be observed for the 125 Hz signal condition than for 250 and 500 Hz, (3) MLDs obtained through monaural testing were insufficient to account for BMLD, and (4) NH listeners using an acoustic CI simulation performed better than CI users, particularly with the N0Sπ condition.

BMLDs and the related phenomenon of noise squelch in bilateral CI users have been reported in just a handful of studies (Nopp *et al.*, 2004; Laszig *et al.*, 2004; van Hoesel, 2004; Long *et al.* 2006; Buss *et al.*, 2008; van Hoesel *et al.*, 2008), and of these, only a few groups have investigated BMLDs using precisely controlled stimulation via single pairs of pitch-matched electrodes that are temporally synchronized through a research interface (van Hoesel, 2004; Long *et al.*, 2006; van Hoesel *et al.*, 2008) as opposed to using the patient's own pair of clinical processors (Laszig *et al.*, 2004; Schleich *et al.*, 2004; Buss *et al.*, 2008). The average BMLD obtained in this study, 4.6 ± 4.9 dB falls

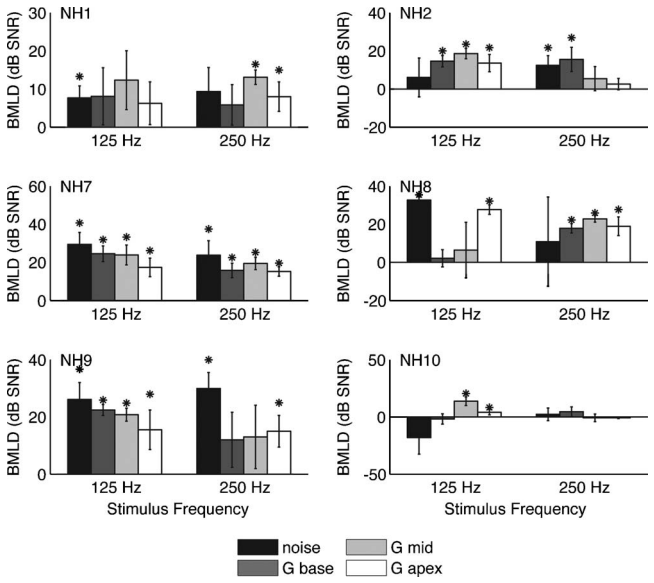


FIG. 8. BMLD for NH listeners subjected to acoustic CI simulation. Format is the same as Fig. 2.

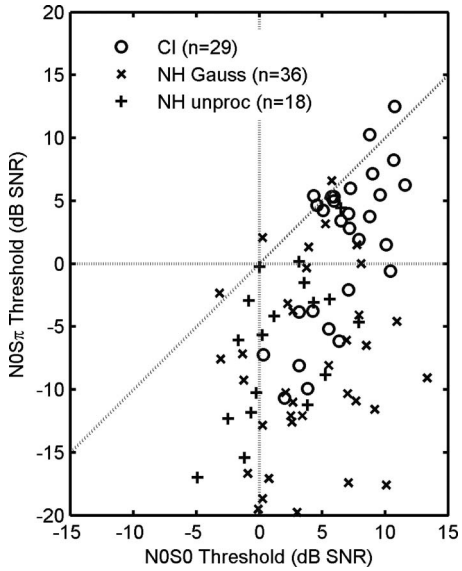


FIG. 9. Comparison of CI users to NH listeners with acoustic CI simulation. N0S0 thresholds are on the *x*-axis. N0Sπ thresholds are on the *y*-axis. Data points are from the 125 and 250 Hz stimulus conditions only. "O's" are data from the five CI subjects. (Mean [N0S0; N0Sπ]=[6.6 ± 2.8; 1.6 ± 6.1] dB SNR.) "X's" are data from six NH listeners using the Gaussian-enveloped tone vocoder. (Mean [N0S0; N0Sπ]=[4.0 ± 4.1; −8.6 ± 7.2] dB SNR.) "+'s" are data from six NH listeners with unprocessed stimuli. (Mean [N0S0; N0Sπ]=[1.6±3.5; −6.3±5.8] dB SNR.)

within the range of BMLDs (1.5–9 dB) reported in literature for bilateral CI users (van Hoesel, 2004; Long *et al.*, 2006; van Hoesel *et al.*, 2008).

Although the stimuli used in this experiment followed that of Long *et al.* (2006), which reported an average BMLD of 9 dB, the data from this study's subject pool displayed considerably more variability. Closer inspection of the data in Long *et al.*, 2006 reveals that average thresholds (as opposed to BMLDs) were higher in this study (Fig. 9). N0S0 threshold was 6.6 dB compared to 3 dB in Long *et al.*, 2006, and N0Sπ threshold was 1.6 dB compared to −6 dB. Individually, data from CI1 and CI5 were most similar to those in Long *et al.*, 2006 where three of four subjects had N0S0 thresholds greater than 0 dB and N0Sπ thresholds near −9 dB. Their consistent performance over most electrode and signal frequency conditions suggests that binaural unmasking can be a useful mechanism for signal detection in noise. The smaller BMLD values from the other subjects, CI2–CI4, contributed to the lack of statistical significance ($p=0.09$) of BMLD over all stimulation conditions and underscore the fact that binaural unmasking may be of limited or no benefit to some bilateral CI users.

In comparing BMLDs with tones and speech, one methodological consideration is that, rather than using an out-of-phase signal as was done in both Long *et al.*, 2006 and the present study, the speech stimuli in van Hoesel *et al.*, 2008 were delayed by 700 μs, which may have contributed to a smaller unmasking effect. For a 125 Hz tone, 700 μs corresponds to a 31.5° phase shift. An out-of-phase signal maximizes the difference between two ears since $\sin(x)-\sin(x+180°)\leq 2$. In addition, since BMLDs for higher frequencies (250, 500 Hz) in this study were not statistically significant, having those higher frequencies present in speech would may not result in observable BMLDs for bilateral CI users. It remains unclear exactly how much contribution 700 μs makes as it is only one of multiple factors that could potentially impact BMLDs for speech.

A. Effect of signal frequency

If the analysis was restricted to just the 125 Hz signal condition (the only tested frequency in Long *et al.*, 2006), the effect of signal phase was statistically significant ($p=0.033$). Because the higher signal frequencies were not tested in Long *et al.*, 2006 it is unknown how their subjects would have performed at 250 and 500 Hz, although it could be speculated that BMLD levels may be similar to CI1 and CI5. With CI1 and CI5, BMLDs were observed for most electrode pair and signal frequency combinations, while two others had BMLDs for only a few combinations, which included 125 Hz. This suggests that the lower signal frequency may be more useful to bilateral CI users as a group, while some individuals could experience binaural unmasking at higher signal frequencies.

The limited BMLD at higher signal frequencies may be related to synchronous activation over broad regions at 300 Hz compared to lower rates (van Hoesel *et al.*, 2009), or poor temporal discrimination of signals above 300 Hz (Zeng, 2002). The same temporal limitations apply even when the stimuli are presented diotically (Carlyon *et al.*, 2008). Although the auditory nerves show highly synchronized responses to electrical stimulation (Hartmann *et al.*, 1984), it is unclear where this temporal limitation is generated along the auditory pathway. If the temporal features are lost before the CI user's auditory system can process the binaural signal, smaller BMLDs could result. In comparison, NH listeners have BMLDs near 10 dB at frequencies as high as 4 kHz (van de Par and Kohlrausch, 1997) as well as larger BMLDs for speech (Levitt and Rabiner, 1967). As speech has a broad spectral composition, the smaller BMLDs of CI users at higher signal frequencies may also help to explain smaller speech BMLDs.

B. Effect of electrode position

BMLDs were also analyzed by electrode position to look for any systematic changes with cochlear region. Although there was a significant effect of cochlear region on thresholds, there was no significant interaction between phase and electrode on thresholds for the 125 Hz condition, meaning that BMLD was not statistically different between electrodes. This indicated that cochlear region was likely not a limiting factor for speech BMLD. The data presented here showed that in at least two subjects, BMLDs were observed at all basal, middle, and apical pairs and could be 10 dB or higher. Since there was no significant effect of electrode position on BMLD, it follows that binaural unmasking was not limited by cochlear region. This result is consistent with NH listeners exhibiting binaural unmasking to the signal in envelope of a high frequency carrier (van de Par and Kohlrausch, 1997).

C. Peripheral mechanisms affecting BMLD

As demonstrated by the present data and prior studies, binaural unmasking can be observed in bilateral CI users, but even the best CI performers in these studies had BMLD values that fell below the 15–20 dB reported for NH listeners (van de Par and Kohlrausch, 1997). This could be due to either inadequate delivery of binaural cues to the peripheral nervous system, inadequate processing by the central nervous system, or a combination of both. It is thus important to consider what underlying mechanisms could be responsible for this performance gap.

The first possibility is that acoustic cues in the envelope are not sufficient to convey binaural information about N0S0 and N0Sπ. However, this is unlikely since NH listeners using the CI simulation had thresholds that were nearly the same as in the unprocessed condition (see Fig. 9). In NH listeners, BMLDs for both the unprocessed and transposed tones were nearly the same, but a reduction of about 6 dB for the transposed tone was noted for a 250 Hz signal (van de Par and Kohlrausch, 1997). In order for the results to be comparable, there must be sufficient acoustic cues remaining in the envelope since fine timing information is discarded by the CI simulation. Despite the same envelope extraction of the signal, actual CI users had higher thresholds than simulated CI, particularly with N0Sπ. For the unprocessed condition, the NH group had BMLD values lower than in litera-

ture. Some of this may be due to the minimal training and inexperience of the subjects, compared to those in van de Par and Kohlrausch, 1997, most of whom were reported to have experience in masking experiments.

A second possible mechanism is that the CI processor cannot adequately convey the acoustic cues in the envelope in time or amplitude. Under these experimental conditions, the pulse rate was set at 1000 and 2000 pps, which was fast enough to represent the signal frequencies used in this study. At high pulse rates (800 pps) ITD sensitivity to the pulses is very poor (van Hoesel and Tyler, 2003), and although this may contribute to some drop in performance, the cues in this study were all based on the envelope rather than in the interaural pulse timing.

A third possibility related to mechanism lies in the electrode-neural interface, such as the availability of suitably pitch-matched electrodes and effects of channel interactions. Because electrode array insertions are rarely identical, electrically elicited pitches are often offset between the two sides. The CI users tested here, and CI1, in particular, commented that when alternately switching between their own left and right processors they perceived there to be a noticeable pitch difference. Although pitch-matching was not directly addressed by the experiments in the present study, it has been shown psychophysically (Henning, 1974; Nuetzel and Hafter, 1981) and physiologically (e.g., Blanks *et al.*, 2007) that the binaural system is most effective when the frequency of sounds between the left and right are nearly the same. In fact, in normal hearing listeners, ITD judgments deteriorate with mismatch in carrier frequency (Henning, 1974; Nuetzel and Hafter, 1981). For CI users, pitch-matched pairs of electrodes are necessary for ITD sensitivity (van Hoesel and Clark, 1997; Lawson *et al.*, 1998; Long *et al.*, 2003) and BMLD (Long *et al.*, 2007).

Since it appeared that envelope was sufficient for transmitting cues for BMLD in this study, we examined whether there were others cues that were contributing to BMLD values. Despite the poor ability to discriminate modulation frequencies (Zeng, 2002), CI users can detect temporal fluctuations occurring up to 4000 Hz (Shannon, 1992). Measuring monaural performance can help to reveal the contribution of binaural processing. To be able to detect the signal in noise under these conditions puts more emphasis on temporal discrimination since a single channel was used and no binaural cues were available. Because MLDs were not significant in this study, BMLDs measured here likely result from true binaural hearing rather than independent and monaural discrimination of envelopes by each ear. Only when both ears were used was there a significant difference in threshold. Since the noise masker was randomly generated for each trial, there is no practical difference between NmSm and NmS-m.

D. Central auditory mechanisms affecting BMLD: Auditory deprivation

The issues discussed so far have been related to technical limitations of the cochlear implant system in processing and delivering binaural stimuli rather than the central auditory mechanisms that underlie binaural sensitivity. Even if normal peripheral processing could be restored in CI users, the central auditory system can still fail to effectively process binaural information. Both the limitations of CI users to discriminate high modulation frequencies (Zeng, 2002) and poor ITD sensitivity (van Hoesel *et al.*, 2002; van Hoesel, 2007; Laback and Majdak, 2008; Litovsky *et al.*, 2010) are examples where central auditory processing showed deficits despite temporally precise delivery of electrical stimuli to the auditory periphery.

In comparing NH listeners using simulations to actual CI users, there are two assumptions to be made about the simulation and the comparison. First is that the vocoder is an accurate simulation of what CI users hear. Gaussian-enveloped tones have not been used in the context of a vocoder, but they have been used for psychophysical, including binaural, experiments (Buell and Hafter, 1988; van den Brink and Houtgast, 1990). Although it would be possible to conduct a similar experiment with a one-channel noise-excited vocoder using band-limited noise, the Gaussian-enveloped tone vocoder may be a better choice. With a noise-excited vocoder, the carrier is noise. There may be interactions between noise in the carrier and any masking noise present in the envelope, particularly in the N0S0 condition. That is, the noise in the carrier may contribute to noise in the envelope such that for normal hearing listeners using the noise vocoder, this becomes a more difficult task without contributing additional insight to BMLD performance. The Gaussian tone pulses do not have this issue as the fine-structures of the tones are controlled and the amplitudes at each time point are defined algorithmically and not influenced by stochastic fluctuations in a noise carrier. For comparison, the N0S0 thresholds were higher for the noise vocoder than it was for the Gaussian-enveloped tone. Although the noise vocoder was implemented with broadband noise, the problem using a band-limited signal to simulate a specific channel remains the same since there is still an issue of a noise carrier interacting with masking noise in the envelope.

If the validity of the vocoder as a simulation is accepted, then it can be assumed that the peripheral auditory systems of NH and CI are receiving equivalent inputs. Therefore, performance differences between the two groups will depend on their ability to process binaural information. With an acoustic CI simulation, differences between NH and CI not explained by pitch-matching and channel interaction may point toward the central auditory system. As a population, the thresholds of CI users lagged behind NH listeners by nearly 10 dB for N0Sπ and but only 2.4 dB for N0S0. This suggests that the binaural performance difference may be in part due to physiological limitations of processing binaural information in the central auditory system. BMLDs for unprocessed stimuli reported here are smaller than those reported in literature (van de Par and Kohlrausch, 1997). Individually, at least one subject (NH4) had a BMLD of 13.9 dB for the unprocessed stimuli. The lower scores may be attributable to the inexperience of the test subjects (most were undergraduate students) compared to the more experienced subjects in van de Par and Kohlrausch, 1997, all but one of whom were indicated as laboratory colleagues.

Examining the clinical data, the best performers in this

study, CI1 and CI5, had the longest experience with bilateral CI use and had prior experience with research testing. Their better performance would be consistent with having a longer recovery of binaural function (Laszig *et al.*, 2004; Buss *et al.*, 2008). The small BMLD values of CI2 may be related to the early onset of hearing impairment. CI3 had late onset of hearing loss, and only 1 year less bilateral experience than CI1 and CI5. Based on this subject's clinical history larger BMLDs would have been expected. CI4 had 3 years of bilateral CI experience. There was some difficulty in finding suitably pitch-matched electrode pairs as noted by the right electrode being matched to apical and middle electrodes on the left over the course of the testing. Although the clinical data do appear to be on first inspection consistent with the BMLD levels of the subjects, due to the small sample size, the various causes of deafness, hearing aid history, and duration of implant use, it would be difficult to predict BMLD performance with any confidence from duration of deafness from the clinical data.

E. Future directions

Despite the large variability between subjects, BMLDs from these single pairs of electrodes (4.6 dB average, and ~9 dB for the top two performers) were still larger than BMLDs for speech (<1.5 dB) reported in CI. The data presented here thus do not fully account for this performance gap. One direction that should be taken is to evaluate BMLDs with multiple and simultaneous masking electrodes. Would the binaural cues presented in one pair of electrode be sufficiently degraded by adjacent masking electrodes such that BMLDs are reduced? Addressing this issue would help to answer to what degree channel interactions affect binaural hearing and provide some clue as to what other factors are limiting BMLD in CI.

V. CONCLUSION

BMLDs were measured in five bilateral CI subjects using the Spear3 research processor. Stimuli were delivered though pitch-matched and loudness-balanced electrode pairs. The cochlear position of the electrode pair was varied as was the signal frequency. Normal hearing listeners were tested with unprocessed and acoustic CI simulations. The conclusions are as follows.

(1) Average BMLD over all stimulus conditions was 4.6 ± 4.9 dB. Significant BMLDs were observed in at least one electrode pair and signal frequency condition for four out of five subjects, with BMLDs varying no more than 3–5 dB over electrode pairs and stimulus conditions. Performance variability was high between subjects and average BMLD values per subject ranged from 1.2 to 10.7 dB;
(2) For the subject group in this study, the effects of phase and electrode pair were significant when considering only the 125 Hz signal frequency condition, as in a prior study, suggesting that BMLDs are more likely to be observed for lower signal frequencies.
(3) Monaurally detectable differences in NmSm and NmS-m were not sufficient to account for BMLDs.
(4) NH listeners had similar unprocessed and simulated CI thresholds for N0S0 and N0Sπ. Actual CI users had higher thresholds than NH with simulated CI, but the difference was more pronounced for N0Sπ, implying a central binaural processing deficit.

The present results support prior studies that demonstrate binaural processing in CI users. However, a performance gap in BMLD still exists between even the best CI users in this study and NH listeners and may be related to a peripheral mechanism such as pitch mismatch, a central mechanism in a deprived auditory system, or a combination of both. The differences in actual and simulated CI performances point to a deficit in true binaural hearing that may have resulted from long-term deprivation of the auditory system.

ACKNOWLEDGMENTS

The authors would like to thank J. Yuan, M. Camilon, and P. Lin for assistance with data collection. This work was supported by NIH Grant Nos. 5 R01 DC003083 (R.L.) and P30 DC008369 (F.-G.Z.).

Blanks, D. A., Roberts, J. M., Buss, E., Hall, J. W., and Fitzpatrick, D. C. (**2007**). "Neural and behavioral sensitivity to interaural time differences using amplitude modulated tones with mismatched carrier frequencies," J. Assoc. Res. Otolaryngol. **8**, 393–408.
Buell, T. N., and Hafter, E. R. (**1988**). "Discrimination of interaural differences of time in the envelopes of high-frequency signal: Integration times," J. Acoust. Soc. Am. **84**, 2063–2066.
Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. H., Clark, M. S., Haynes, D. S., Labadie, R. F., Amberg, S., Roland, P. S., Kruger, P., Novak, M. A., Wirth, J. A., Black, J. M., Peters, R., Lake, J., Wackym, P. A., Firszt, J. B., Wilson, B. S., Lawson, D. T., Schatzer, R., D'Haese, P. S., and Barco, A. L. (**2008**). "Multicenter U.S. bilateral MED-EL cochlear implantation study: Speech perception over the first year of use," Ear Hear. **29**, 20–32.
Carlyon, R. P., Long, C. J., and Deeks, J. M. (**2008**). "Pulse-rate discrimination by cochlear-implant and normal-hearing listeners with and without binaural cues," J. Acoust. Soc. Am. **123**, 2276–2286.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Gray, L., Kesser, B., and Cole, E. (**2009**). "Understanding speech in noise after correction of congenital unilateral aural atresia: Effects of age in the emergence of binaural squelch but not in use of head-shadow," Int. J. Pediatr. Otorhinolaryngol. **73**, 1281–1287.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Hall, J. W., Buss, E., Grose, J. H., and Dev, M. B. (**2004**). "Developmental effects in the masking-level difference," J. Speech Lang. Hear. Res. **47**, 13–20.
Hall, J. W., III, and Grose, J. H. (**1993**). "Short-term and long-term effects on the masking level difference following middle ear surgery," J. Am. Acad. Audiol. **4**, 307–312.
Hartmann, R., Topp, G., and Klinke, R. (**1984**). "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea," Hear. Res. **13**, 47–62.
Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.
Hinojosa, R., and Lindsay, J. R. (**1980**). "Profound deafness. Associated sensory and neural degeneration," Arch. Otolaryngol. **106**, 193–209.
Hirsh, I. J. (**1948**). "Binaural summation and interaural inhibition as a function of the level of masking noise," Am. J. Psychol. **61**, 205–213.
Hogan, S. C., Meyer, S. E., and Moore, D. R. (**1996**). "Binaural unmasking returns to normal in teenagers who had otitis media in infancy," Audiol. Neuro-Otol. **1**, 104–111.
Laback, B., and Majdak, P. (**2008**). "Binaural jitter improves interaural time-

difference sensitivity of cochlear implantees at high pulse rates," Proc. Natl. Acad. Sci. U.S.A. **105**, 814–817.
Laszig, R., Aschendorff, A., Stecker, M., Muller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T., Battmer, R. D., Bohm, M., Steffens, T., Strutz, J., Linder, T., Probst, R., Allum, J., Westhofen, M., and Doering, W. (**2004**). "Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results," Otol. Neurotol. **25**, 958–968.
Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C., Jr., McElveen, J. T., Jr., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Levitt, H., and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–829.
Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (**2006a**). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," Ear Hear. **27**, 714–731.
Litovsky, R. Y. (**1997**). "Developmental changes in the precedence effect: Estimates of minimum audible angle," J. Acoust. Soc. Am. **102**, 1739–1745.
Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., and Lake, J. (**2006b**). "Bilateral cochlear implants in children: Localization acuity measured with minimum audible angle," Ear Hear. **27**, 43–59.
Litovsky, R. Y., Jones, G. L., Agrawal, S., and van Hoesel, R. (**2010**). "Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans," J. Acoust. Soc. Am. **127**, 400–414.
Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.
Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Yu, G. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.
Long, C. J., Carlyon, R. P., and Litovsky, R. Y. (**2007**). "Binaural unmasking with 'transposed' stimuli in bilateral cochlear implant users," Conference on Implantable Auditory Prosthesis, Granlibakken, Tahoe City, CA (abstract).
Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (**2006**). "Binaural unmasking with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **7**, 352–360.
Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.
Lu, T., Carroll, J., and Zeng, F.-G. (**2007**). "On acoustic simulations of cochlear implants," Conference on Implantable Auditory Prostheses (abstract), Lake Tahoe, CA, July 15–20.
Müller, J., Schon, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," Ear Hear. **23**, 198–206.
Neuman, A. C., Haravon, A., Sislian, N., and Waltzman, S. B. (**2007**). "Sound-direction identification with bilateral cochlear implants," Ear Hear. **28**, 73–82.
Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants," Ear Hear. **25**, 205–214.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant," Ear Hear. **25**, 197–204.
Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.
Silman, S., Gelfand, S. A., and Silverman, C. A. (**1984**). "Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids," J. Acoust. Soc. Am. **76**, 1357–1362.
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.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Tyler, R. S., and Summerfield, A. Q. (**1996**). "Cochlear implantation: Relationships with research on auditory deprivation and acclimatization," Ear Hear. **17**, 38S–50S.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
van den Brink, W. A., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.
Van Deun, L., van Wieringen, A., Francart, T., Scherf, F., Dhooge, I. J., Deggouj, N., Desloovere, C., Van de Heyning, P. H., Offeciers, F. E., De Raeve, L., and Wouters, J. (**2009b**). "Bilateral cochlear implants in children: Binaural unmasking," Audiol. Neuro-Otol. **14**, 240–247.
Van Deun, L., van Wieringen, A., Van den Bogaert, T., Scherf, F., Offeciers, F. E., Van de Heyning, P. H., Desloovere, C., Dhooge, I. J., Deggouj, N., De Raeve, L., and Wouters, J. (**2009a**). "Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing," Ear Hear. **30**, 178–190.
van Hoesel, R. J. (**2007**) "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.
van Hoesel, R., Bohm, M., Pesch, J., Vandali, A., Battmer, R. D., and Lenarz, T. (**2008**). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies," J. Acoust. Soc. Am. **123**, 2249–2263.
van Hoesel, R. J. M., Jones, G. L., and Litovsky, R. Y. (**2009**). "Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse-rate, modulation-rate, and place of stimulation," J. Assoc. Res. Otolaryngol. **10**, 557–567.
van Hoesel, R., Ramsden, R., and Odriscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.
van Hoesel, R. J. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.
van Hoesel, R. J., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.
van Hoesel, R. J., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). "Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject," J. Acoust. Soc. Am. **94**, 3178–3189.
van Hoesel, R. J., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
Verschuur, C. A., Lutman, M. E., Ramsden, R., Greenham, P., and O'Driscoll, M. (**2005**). "Auditory localization abilities in bilateral cochlear implant recipients," Otol. Neurotol. **26**, 965–971.
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) **352**, 236–238.
Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear Res. **174**, 101–106.
Zeng, F. G., and Galvin, J. J., III (**1999**). "Amplitude mapping and phoneme recognition in cochlear implant listeners," Ear Hear. **20**, 60–74.

Multicenter evaluation of signal enhancement algorithms for hearing aids[a)]

Heleen Luts,[b)] Koen Eneman,[c)] and Jan Wouters
ExpORL, Department of Neurosciences, Katholieke Universiteit Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium

Michael Schulte and Matthias Vormann
Hörzentrum Oldenburg GmbH, D-26129 Oldenburg, Germany

Michael Buechler and Norbert Dillier
Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Zürich, CH-8091 Zürich, Switzerland

Rolph Houben and Wouter A. Dreschler
AMC, KNO-Audiologie, 1105 AZ Amsterdam, The Netherlands

Matthias Froehlich and Henning Puder
Siemens Audiologische Technik GmbH, D-91058 Erlangen, Germany

Giso Grimm and Volker Hohmann
Medical Physics Section, Carl von Ossietzky-Universität Oldenburg, D-26111 Oldenburg, Germany

Arne Leijon
Sound and Image Processing Laboratory, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden

Anthony Lombard
Multimedia Communications and Signal Processing, Universität Erlangen-Nürnberg, D-91058 Erlangen, Germany

Dirk Mauler
Institute of Communication Acoustics (IKA), Ruhr-Universität Bochum, IC 1/132, D-44780 Bochum, Germany

Ann Spriet
ESAT/SCD-SISTA, Katholieke Universiteit Leuven, B-3001 Leuven-Heverlee, Belgium

(Received 23 February 2009; revised 6 January 2010; accepted 6 January 2010)

In the framework of the European HearCom project, promising signal enhancement algorithms were developed and evaluated for future use in hearing instruments. To assess the algorithms' performance, five of the algorithms were selected and implemented on a common real-time hardware/software platform. Four test centers in Belgium, The Netherlands, Germany, and Switzerland perceptually evaluated the algorithms. Listening tests were performed with large numbers of normal-hearing and hearing-impaired subjects. Three perceptual measures were used: speech reception threshold (SRT), listening effort scaling, and preference rating. Tests were carried out in two types of rooms. Speech was presented in multitalker babble arriving from one or three loudspeakers. In a pseudo-diffuse noise scenario, only one algorithm, the spatially preprocessed speech-distortion-weighted multi-channel Wiener filtering, provided a SRT improvement relative to the unprocessed condition. Despite the general lack of improvement in SRT, some algorithms were preferred over the unprocessed condition at all tested signal-to-noise ratios (SNRs). These effects were found across different subject groups and test sites. The listening effort scores were less consistent over test sites. For the algorithms that did not affect speech intelligibility, a reduction in listening effort was observed at 0 dB SNR.
 [DOI: 10.1121/1.3299168]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Gv, 43.60.Fg [BCM] Pages: 1491–1505

[a)]Portions of this work were presented at Acoustics'08, Paris, France, June 2008; International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, August 2008; the 16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland, September 2008; and the 9th European Federation of Audiology Societies Congress, Tenerife, Spain, June 2009.

[b)]Author to whom correspondence should be addressed. Electronic mail: heleen.luts@med.kuleuven.be
[c)]Present address: Group T-International University College Leuven, Andreas Vesaliusstraat 13, 3000 Leuven, Belgium.

I. INTRODUCTION

The main complaints of hearing aid users are problems with speech understanding in noisy listening environments and reduced spatial awareness (Kochkin, 2005; Noble and Gatehouse, 2006). With the advent of fully digital hearing aids around 1995, there was a general belief that new signal processing strategies would be developed and (older) schemes would be integrated as software modules in the new digital engines. Several speech-in-noise enhancement strategies, including directional microphone systems, can be considered as a front-end to the core processing of the hearing aid or cochlear implant. Whereas a lot of research has been carried out on speech enhancement approaches in challenging noisy environments, until now only a limited number of algorithms have been implemented in commercial devices (Dillon, 2001; Blamey, 2005; Spriet *et al.*, 2007; Hu and Loizou, 2008). The implementation of a signal processing scheme in a hearing aid or cochlear implant makes strong demands in terms of computational complexity and processing delay, and requires extensive validation through physical and perceptual validation tests.

The reason for the limited application of digital signal enhancement techniques in commercial hearing instruments is twofold. First, few signal processing schemes have been developed and sufficiently evaluated for real-world application. Second, only a limited number of these signal processing schemes have actually demonstrated real benefits for hearing aid users both in laboratory environments and under daily listening conditions. Moreover, the results are hard to compare across different developers and test sites. Based on results published in the engineering literature, replication is sometimes difficult because of incomplete knowledge of details. In addition, the physical evaluations and objective measures are not always relevant to the perception of speech in noise. For instance, not only is the signal-to-noise ratio (SNR) usually important, but also perception-related measures such as the intelligibility-weighted SNR (Greenberg *et al.*, 1993) and the speech intelligibility index (SII) (ANSI S3.5-1997, 1997). The evaluation should not be limited to experiments with artificial noise (e.g., white noise), but should include extensive testing with real-world signals as well. Additionally, perceptual evaluation experiments should be included, with both normal-hearing (NH) and hearing-impaired (HI) subjects, i.e., representative of the target population.

One of the subprojects within the framework of the European research project HearCom (Hearing in the Communication Society) has focused on the development and evaluation of signal enhancement techniques for improving speech understanding by use of hearing aids. Based on physical performance measures (Eneman *et al.*, 2008a, 2008b), five signal enhancement techniques (single- as well as multi-channel) were selected from a large set of state-of-the-art algorithms and implemented on a common real-time low-delay test platform (Grimm *et al.*, 2006). The study presented in this paper comprises the perceptual evaluation of these five algorithms at four different test sites with large numbers of normal-hearing listeners and hearing aid users with flat or sloping moderate hearing losses.

Adaptive speech reception threshold (SRT) tests were performed to evaluate the effect of the signal enhancement algorithms on speech intelligibility. Several previous studies on the perceptual benefits of single-channel noise reduction algorithms described increased comfort or ease of listening, while finding no evidence for improvement in speech intelligibility (Marzinzik and Kollmeier, 1999; Walden *et al.*, 2000; Ricketts and Hornsby, 2005; Bentler *et al.*, 2008). Therefore, in the current study, listening effort and overall preference were also assessed. The listening effort scaling (LES) and preference rating (PR) measurements were carried out at a number of fixed SNRs up to +10 dB, which are representative of many daily listening conditions.

The following research questions are addressed and discussed in this paper: Do signal processing strategies lead to similar outcomes across different test sites, given the different test environments/rooms, test materials, and evaluation test platforms? Do the enhancement strategies yield different outcomes for different subject groups? How do the results for both objective and subjective perceptual measures vary across the signal processing algorithms?

II. SIGNAL ENHANCEMENT TECHNIQUES

This section describes the five signal enhancement algorithms that were selected and evaluated.

A. Single-channel noise suppression based on perceptually optimized spectral subtraction (SC1)

Spectral subtraction is a well-known computationally-efficient noise reduction technique, based on the idea of subtracting an estimate of the noise spectrum from an estimate of the spectrum of the noisy signal (e.g., Boll, 1979). The present implementation of this principle is a variant known as magnitude spectral subtraction. To reduce the well-known phenomenon of musical noise artifacts, the present version was perceptually optimized to control the trade-off between speech distortion and noise suppression. The input signal is analyzed in 50% overlapping Hann-windowed blocks of duration 4 ms, with 2-ms frame update steps. The enhanced speech discrete Fourier transform (DFT) coefficients $\hat{S}(m,k)$, with frame index m and DFT bin k, are obtained as $\hat{S}(m,k) = H_{SS}(m,k) \cdot Y(m,k)$, where $Y(m,k) = S(m,k) + N(m,k)$ are the DFT coefficients of the input speech $S(m,k)$ and noise $N(m,k)$. $H_{SS}(m,k)$ is a real-valued time- and frequency-dependent adaptive filter gain, calculated as

$$H_{SS}(m,k) = qH_{SS}(m-1,k) + (1-q)\left(1 - \alpha(m,k)\frac{\sqrt{|\hat{N}(m,k)|^2}}{|Y(m,k)|}\right),$$

and limited to ensure that $H_{SS}(m,k) \geq 0$. Here, the first-order recursive low-pass smoothing parameter is $q = 0.9$, giving an effective time constant of 19 ms for the adaptive gain variations. $|\hat{N}(m,k)|^2$ represents the noise power spectrum as esti-

mated from the modulation pattern of the noisy input signal, using the minimum statistics algorithm (Martin, 2001) based on a memory buffer of the most recent 1.5 s of the input signal. The degree of spectral subtraction is controlled by the frequency-dependent "aggressiveness factor" $\alpha(m,k)$, with values between 0 and 1.6, determined as a function of the estimated frequency-dependent noisy-signal-to-noise ratio (NSNR) $z(m,k)=|Y(m,k)|^2/|\hat{N}(m,k)|^2$, as

$$\alpha(m,k) = a - a\ \tanh[b(10\ \log_{10}\ z(m,k) - c)^2].$$

The parameters a, b, and c have been manually tuned to optimize the perceptual performance of the algorithm. Their values are adapted slowly, depending on the overall (not frequency-dependent) SNR, calculated using first-order recursive low-pass filter smoothing of the estimated speech and noise frame power values, with time constants of 1.6 s. For low overall SNRs, of 0 dB or less, the maximum aggressiveness is 0.6 for $z(m,k)$ corresponding to 3 dB. For SNRs of 20 dB or higher, the maximum aggressiveness is 1.6 for $z(m,k)$ corresponding to 7.5 dB. The aggressiveness factor is zero for any $z(m,k)$ below −15 dB and above +25 dB. Finally, the enhanced time-domain output signal is reconstructed from $\hat{S}(m,k)$ by the inverse discrete Fourier transform (IDFT) and overlap-add operations.

B. Wiener-filter-based single-channel noise suppression (SC2)

A second single-channel noise suppression algorithm was based on the Wiener filter, which minimizes the mean-squared error between the (unknown) desired speech signal and a filtered version of the observed noisy speech. Since speech is stationary only over short time intervals, statistical expectation operations have to be replaced with short-term averages. Therefore, instead of using the actual a priori SNR, estimated a priori SNR values were computed as described by Ephraim and Malah (1984). The samples of the observed noisy speech signal were partitioned into frames of 8-ms length, overlapping by 6 ms. The frame shift was 2 ms. Each frame of data was weighted with the square root of a Hann window and was then transformed to the DFT domain. The enhanced speech spectral coefficients $\hat{S}(m,k)$, with frame index m and DFT bin k, were obtained as $\hat{S}(m,k) = H(m,k)\cdot Y(m,k)$, where $Y(m,k)=S(m,k)+N(m,k)$ are the DFT coefficients of speech $S(m,k)$ and noise $N(m,k)$. $H(m,k)$ is a time- and frequency-dependent gain. For the Wiener filter approach, $H(m,k)$ is real-valued. Only the amplitudes of the noisy DFT coefficient are changed and the phases are left unchanged. The Wiener filter solution

$$H(m,k) = \frac{E\{|S(m,k)|^2\}}{E\{|S(m,k)|^2\} + E\{|N(m,k)|^2\}}$$

is obtained by a minimization of the mean-square error between estimated and true speech spectral coefficients (e.g., Vary and Martin, 2006). $E\{X\}$ denotes the expectation of variable X. After IDFT and weighting the data again with the square root of a Hann window, the time-domain signal was reconstructed via overlap-add operations. The estimation of the noise power spectral density $E\{|N(m,k)|^2\}$ via minimum statistics (Martin, 2001) rests on the observation that the power of the noisy speech signal frequently decays to the level of the noise. An estimate of the noise power can hence be obtained by tracking minima of the spectral power. Then, the bias between minimum and mean is compensated for. Due to the minimum principle, noise power estimation via minimum statistics does not require explicit voice activity detection.

The algorithm variant used was a complexity- and delay-optimized solution. The reduced frame length and frame shift result in a larger frame overlap and hence a significantly increased correlation of the spectral data of successive frames. In Mauler and Martin (2006), it was shown that highly correlated spectral data require an adjustment of the noise power bias correction in the minimum statistics algorithm. Alternatively, without modification of this bias correction, the correlation of spectral data can be reduced if its estimation is based on subsampled frames (i.e., less overlapped). In the algorithm under test, the noise power estimate was updated only after three frame shifts, i.e., after every 6 ms, while the enhanced speech spectrum was computed after every frame shift.

C. Broadband blind source separation (BSS) based on second-order statistics

The aim of blind source separation is to recover the original source signals from an observed set of signal mixtures, solely based on the assumption of statistical independence of the original source signals. The term "blind" implies that the mixing process and the original source signals are unknown. In acoustical scenarios, like in a hearing aid application, signals propagate through multiple paths due to acoustical reflections. The microphone signals result therefore from a superposition of multiple scaled and delayed versions of the original source signals. This superposition of sounds emitted from the same source can be described as a convolutional product. Hence, in a multiple-source scenario, the source signals are mixed in a convolutive manner. Separation is achieved by applying a set of adaptive filters to the microphone signals. The adaptation of the BSS filters is performed by forcing the BSS outputs to become statistically independent, as described in Buchner *et al.* (2004, 2005a) and Aichner *et al.* (2006), where a class of broadband BSS algorithms was derived. Broadband BSS approaches process all frequency bins jointly, in contrast to the so-called narrowband approaches, which consider each frequency bin independently.

The (broadband) BSS algorithm selected here is a computationally efficient, low-delay variant using frequency-domain-based fast convolution techniques and second-order statistics (see Aichner *et al.*, 2006). The BSS algorithm was applied to binaural hearing aids, using the front microphone signal from each hearing aid as its inputs. This two-microphone implementation allows the separation of two point sources and offers two output signals. Thus, the output containing the desired signal had to be selected and presented to the hearing aid user.

Based on the approach described by Buchner *et al.* (2005b), the time-difference-of-arrival (TDOA) of the sound waves originating from the separated sources can be determined from the BSS filters, without any prior knowledge of the microphone positions. Note that, as the microphone spacing is not accurately known, and since head-shadow effects influence the TDOA estimate, accurate direction-of-arrival (DOA) cannot be calculated for each separated source. However, it is assumed here that the desired source is located approximately in front of the hearing aid user. The TDOA estimates are then sufficient for identifying the most frontal source (i.e., the source with the smallest TDOA). The BSS output channel containing the desired source is selected based on this information.

D. Spatially preprocessed speech-distortion-weighted multi-channel Wiener filtering (MWF)

The MWF is an adaptive noise suppression technique that is based on work described in Spriet *et al.*, 2004, 2005 and Doclo *et al.*, 2005, 2007. It consists of a fixed spatial preprocessor, i.e., a fixed beamformer and blocking matrix, and an adaptive stage. The fixed beamformer creates a so-called speech reference by steering a microphone beam toward the front. The blocking matrix creates so-called noise references by steering zeros toward the front so that the noise contributions are dominant compared to the speech leakage contributions. As a consequence, the MWF can be viewed as a variant of the well-known generalized sidelobe canceler (GSC) structure. Whereas in the case of the GSC the filter weights converge to a solution that merely reduces the residual noise, the cost function of the adaptive stage in the MWF approach minimizes a weighted sum of the residual noise energy and the speech-distortion energy. In this way, a trade-off is provided between noise reduction and speech distortion. If the trade-off parameter in the cost function is set to infinity, speech distortion is completely ignored and the algorithm reduces to a GSC structure. The MWF algorithm can therefore be considered as an extension of the GSC. As the MWF approach incorporates a trade-off between noise suppression and speech distortion, the algorithm is more robust against speech leakage than the standard GSC (Spriet *et al.*, 2004). Several algorithm variants have been developed, leading to low-cost implementation and/or improved performance (Spriet *et al.*, 2004, 2005; Doclo *et al.*, 2005, 2007). Here, a three-microphone version of the algorithm was used that relies on a frequency-domain variant of the cost function and that uses efficient updating of the correlation matrix (Doclo *et al.*, 2007).

E. Binaural coherence dereverberation filter (COH)

Dereverberation algorithms are designed to increase listening comfort and speech intelligibility in reverberant environments and diffuse background noise (e.g., babble). The dereverberation technique used here was a binaural-coherence-filtering-based approach that builds on work described by Wittkop and Hohmann (2003). It estimates the coherence, i.e., the signal similarity, between the signals captured at the left and right ears. The estimate is computed in different frequency bands using a fast Fourier transform (FFT)-based filterbank with a non-linear frequency mapping that approximates the Bark scale. As a coherence estimate, the absolute value of the time average of the complex phase difference vector e^(i IPD) is computed. If the signals are coherent in a specific frequency band, the sound is expected to be directional. Hence the gain in the frequency band is set to a high value. If the coherence is low, a diffuse sound field is present, and accordingly, the frequency band is attenuated. The phase in a frequency band is estimated by taking the complex-valued average across frequency of the complex-valued signal spectrum, weighted with a frequency-dependent filter shape function. This filter shape function is an asymmetric Hann window with the maximum value of 1 at the center frequency and the value of 0 at the neighboring band center frequencies. The IPD is calculated in each frequency band by taking the difference of the phases of the respective frequency bands. The IPD is then translated into the complex plane, e^(i IPD), and this complex phasor is filtered with a first-order low-pass with a time constant of 40 ms (Grimm *et al.*, 2009b). The vector strength, i.e., the absolute value of the low-pass filtered complex IPD, is the estimate of the coherence. The frequency-dependent gains are derived from the phase difference vector strength by applying an exponent (between 0.5 and 2) to the coherence estimate. High values of the exponent provide efficient filtering, but lead to more audible artifacts. Because of diffraction effects around the head, the coherence is always high at low frequencies, independent of the type of signal. At medium and high frequencies the coherence is low for diffuse babble noise and for reverberant signal components (late reflections), while it is high for the direct-path contribution of the signal of interest. Due to the temporal and spectral fluctuations of speech, the ratio between non-coherent noise-like signal components and the desired target signal components may vary across time and frequency. Hence, by applying appropriate gains in frequency bands and at times where the non-coherent signal components are dominating, reverberant signal components and diffuse noise can be suppressed relative to direct-path signal components.

III. IMPLEMENTATION ON A COMMON EVALUATION PLATFORM

The signal enhancement algorithms were implemented on a common real-time hardware/software platform, called the personal hearing system (PHS). The hardware platform consists of a (laptop) PC running a real-time low-latency Linux operating system. The PC is equipped with a multichannel RME sound card, which is connected to a pair of hearing aids via a pre-amplifier box. The signals were sampled with 24-bit resolution at 32-kHz sampling rate, resampled to 16 kHz, and the signal processing was performed with single precision floating point resolution (32 bits/sample). With this setup, a glitch-free total delay of approximately 10 ms from the microphones to the receiver was attained. All algorithm developers incorporated a C/C++ implementation of their algorithm into the master hearing aid (MHA). This software environment simulates the processing

TABLE I. Characteristics of the different signal enhancement approaches when implemented on the PHS system.

Algorithm	Monaural/ binaural processing	No. of microphones used	Computational complexity (% CPU time required)	Input/ output delay (ms)
SC1	Monaural	1	8.3	10.8
SC2	Monaural	1	4.2	16.8
BSS	Binaural	2	59.9	10.8
MWF	Monaural	3	4.3	13.2
COH	Binaural	2	1.2	10.6

performed by a hearing aid, and is controlled by the PHS. Hence, apart from passing signals to and from the algorithms, the MHA software is also responsible for applying basic hearing aid processing to the signals, such as frequency-dependent gain setting according to the audiogram of the subject and compensative calibration of microphones and receivers. More details are given later. More information about the MHA and PHS can be found in Grimm *et al.* (2006, 2009a).

The devices used in this study were Siemens Acuris behind-the-ear hearing aids each with three microphones (in end-fire configuration) and a single receiver. There was no processor in the hearing aids themselves. All signal processing was done externally on the PC. The single-channel noise suppression algorithms, SC1 and SC2, only used the front microphone signal, and the same algorithm was used for each aid signal, with identical parameters, independently of each other (double monaural system). The BSS and COH approaches were truly binaural algorithms using the front microphones of the left and right hearing aids as their inputs. The MWF beamformer processed all three microphone signals of each of the hearing aids. The same algorithm was run for the left and right hearing aid signals, independently of each other (for an overview, see Table I).

For possible future implementation in commercial hearing aids, the computational complexity and the input/output delay of the algorithms have to be considered (see Table I). The computational complexity measurements were performed on a Dell Latitude D610 with Intel Pentium M 1.6 GHz processor. The baseline processing of the PHS-MHA system when all signal enhancement algorithms were switched off required 10.3% of CPU time. The total input/output delay from the signal sent into the AD converter to the signal that appears at the DA converter output was measured on a Dell Latitude D620 with Intel Core Duo 1.83 GHz processor. Table I shows the total input/output delay, which includes the combined delay of the PHS-MHA system and of the selected signal enhancement algorithm. With all signal enhancement algorithms switched off an input/output delay of 10.6 ms was measured. Both the PHS-MHA system and the signal enhancement algorithms operate in the frequency domain and thus require an analysis and re-synthesis filterbank. The algorithms SC1 and COH use the analysis and synthesis filterbank of the PHS-MHA system. The current implementations of SC2, BSS, and MWF, however, use a separate filterbank, which causes an additional delay. In an optimized implementation with a shared filterbank, this delay could be reduced. According to Stone and Moore (1999), the auditory effects of hearing aid delays are likely to become disturbing for delays exceeding 20 ms. In the current study, none of the processing delays exceeded 20 ms.

IV. MATERIALS AND METHODS

All five signal enhancement algorithms described in Sec. III were evaluated through listening tests in Dutch and German across four different test sites in Belgium (ExpORL, Department of Neurosciences, K.U.Leuven, "BE"), The Netherlands (AMC, KNO-Audiologie, Amsterdam, "NL"), Germany (Hörzentrum Oldenburg GmbH, "DE"), and Switzerland (Department of Otorhinolaryngology, University Hospital Zürich, "CH").

A. Subjects

In total, 109 subjects participated. Three groups were defined, based on audiogram information only. One group consisted of 38 NH subjects with average hearing thresholds better than or equal to 20 dB hearing level (HL) for octave frequencies between 250 and 8000 Hz. The other 71 subjects had a moderate sensorineural hearing loss and were experienced bilateral hearing aid users (at least 6 months of experience). Bilateral acoustic inputs were used for all tests. Therefore, only a limited amount of asymmetry between the two ears could be tolerated. To this end, the average absolute difference between the left and the right hearing thresholds at the octave frequencies between 500 and 4000 Hz (referred to as symmetry) was intended to be below 10 dB (range 1–19 dB, exceeding 10 dB only in four subjects). The HI subjects were divided into two groups based on the slope of their hearing loss. Slope was defined as the difference between the maximum and minimum hearing thresholds for octave frequencies between 500 and 4000 Hz. The group of hearing-impaired subjects with a flat hearing loss (HI-F) had a slope (averaged for the two ears) of no more than 25 dB. The group with a sloping hearing loss (HI-S) had a slope of more than 25 dB. More details of the subject groups can be found in Table II. The numbers of subjects tested at each test site were 30, 30, 28, and 21 for BE, NL, DE, and CH, respectively.

B. Fitting

Bilateral fittings were based on the audiogram (octave frequencies between 250 and 8000 Hz). The hearing aids were fitted with the NAL-RP prescription rule (Byrne *et al.*, 1991). No compression was included, only a limiter set at 100 dB sound pressure level (SPL) within each frequency band. The gain was adjusted in eight frequency bands, centered at the frequencies 250, 500, 1000, 1500, 2000, 3000, 4000, and 6000 Hz. Fine-tuning was limited to two situations, namely, overall gain experienced as too loud by the subject, or the occurrence of feedback. These situations occurred for approximately 25% and 15% of the HI subjects, respectively. Instead of earmolds, disposable foam earplugs with tubing were used for all ears. For NH subjects, the insertion gain was set to 0 dB.

TABLE II. Description of the subject groups. Degree, slope, and symmetry were calculated for octave frequencies from 500 to 4000 Hz. Slope was defined as the difference between the maximum and minimum hearing thresholds. Symmetry was defined as the average absolute difference between the left and the right hearing thresholds.

		Age (years)		Degree (dB HL)		Slope (dB)		Symmetry (dB)	
Group	*N*	Average	Range	Average	Range	Average	Range	Average	Range
NH	38	30	16–52	4	−3 to 12	9	0–18	4	0–9
HI-F	34	62	22–79	49	34–65	17	5–25	5	1–11
HI-S	37	68	51–80	45	34–61	44	28–70	6	1–19

C. Room characteristics

To investigate the effect of reverberation time on the performance of the algorithms, evaluation tests were conducted in two types of listening room: office-like rooms, representative of many everyday listening conditions, and highly reverberant rooms, which were included to assess the algorithms under more challenging acoustic conditions. At all test sites, the tests were conducted in an office-like room, with a reverberation time (RT_{60}) for frequencies between 300 and 8000 Hz (measured according to the ISO 3382-1997 standard) between 300 and 600 ms. The critical distances for these rooms were 128, 102, 186, and 145 cm for BE, CH, DE, and NL, respectively. Additionally, at two test sites (BE and DE), the algorithms were assessed in a highly reverberant room with an RT_{60} greater than 1 s. The critical distances of these rooms were 37 and 119 cm, for BE and DE, respectively. The maximum background noise level in the rooms was 35 dB(A). For both types of rooms, no additional specifications (room dimensions, room organization, position of test subject, etc.) were defined. At DE the measurements in office-like and reverberant conditions were carried out in a room whose acoustics can be changed systematically.

D. Environmental conditions

During the tests, the listener was given a hearing aid pair controlled by the PHS-MHA. The subject was seated in the test room amidst four loudspeakers that were positioned at 0°, 90°, 180°, and 270°, at 1 m distance from the center of the listener's head. In the highly reverberant room of partner DE, the distance between loudspeakers and listener was 2 m. All loudspeakers were directed toward the listener. Speech was always presented through the front loudspeaker. Two different noise configurations were used. One consisted of three uncorrelated noise sources at 90°, 180°, and 270° (S0N90/180/270). All algorithms were evaluated in this noise scenario, in both room types. In the highly reverberant room, this led to a diffuse noise, as the loudspeakers were positioned outside of the critical distance. In the office-like room, however, this was not the case and the noise field is called pseudo-diffuse. A second noise configuration, which was only used in the office-like room, consisted of one interfering noise source at 90°, i.e., to the listener's right (S0N90). Only two algorithms (BSS and MWF) were evaluated in this single point-source scenario.

A perceptually relevant noise type was chosen for the evaluation of the algorithms. A multitalker babble from the CD *Auditory Tests (Revised)* (Auditec, St. Louis, MO) was used. The noise sources were calibrated to produce a combined sound level of 65 dB(A) at the center of the listener's head (without the listener being present). The speech level was then adjusted to obtain the desired SNR. The noise started 5 s before the first speech sound was presented, to allow the algorithms to initialize properly.

E. Evaluation measures and speech materials

All subjects performed three types of listening tests: adaptive SRT tests, LES, and PR. SRT, LES, and PR were measured for all algorithms (and for the unprocessed condition) at all four test sites in an office-like room with three sources of multitalker babble. The LES and PR tests were carried out after the SRT test so the subjects were already familiar with the listening situation. For the SRT test, additional measurements were performed in other test conditions at a subset of test sites.

1. Adaptive SRT test

At the Dutch-speaking test sites, BE and NL, the open-set VU-sentences (male speaker) were used (Versfeld *et al.*, 2000). Sentence scoring was applied. This speech material consists of 39 lists of 13 sentences. An adaptive one-up one-down test procedure was used. The noise was presented at a fixed level of 65 dB(A) and the level of the target speech was adapted in 2-dB steps. The SRT was defined as the average of the ten last speech presentation levels (including the 14th level, calculated based on the response to the 13th sentence). At BE, the APEX software was used for the SRT testing (Francart *et al.*, 2008). At NL, custom software was used.

At the German-speaking test sites, CH and DE, the closed-set OLSA sentence test was used (Wagener *et al.*, 1999). This speech material (male speaker) consists of 10 lists of 10 sentences with a fixed structure (5 words), combined to give lists of 20 sentences. Sentence scoring was used. The adaptive procedure and the fit of the psychometric curve were as described by Brand and Kollmeier (2002). The OLSA sentence test was used as incorporated in the Oldenburg measurement applications software package developed by Hörtech Oldenburg. Two training lists were used in quiet (20 sentences each).

SRTs using the algorithms were always compared to the SRT for the unprocessed condition, i.e., when all signal enhancement algorithms were switched off and only the basic processing of the PHS-MHA system was activated. In this

TABLE III. Overview table indicating which partners performed which SRT test conditions.

Test condition			Algorithm					
Room	Noise scenario	Masker azimuth	Unpr	SC1	SC2	BSS	MWF	COH
Office-like room	Pseudo-diffuse	90°/180°/270°	All	All	All	All	All	All
Office-like room	Point-source	90°	BE			BE	BE	
			CH			CH	CH	
			NL			NL	NL	
Reverberant room	Diffuse	90°/180°/270°	BE	BE	BE	BE	BE	BE
			DE	DE	DE	DE	DE	DE

way, changes in speech understanding in noise were assessed. An overview of all test conditions is shown in Table III. All conditions were tested in randomized order and conducted twice (test and retest) in one or more test sessions, depending on the total number of conditions to test.

2. LES

Each subject had to rate the listening effort for each algorithm and for the unprocessed condition at five different SNRs: −10, −5, 0, +5, and +10 dB. The LES was performed using a 13 point scale (7 subcategories, with 1 empty button in between). The subcategories ranged from "extreme effort" (score 6) to "no effort" (score 0). LES was performed in the office-like room, with the three interfering noise sources (S0N90/180/270). The same speech material as for the SRT test was used. All conditions were tested and retested, resulting in a total of 60 ratings.

3. PR

As the perceptual differences between the algorithms were expected to be small, a paired-comparison test was used for the PR. Each algorithm was compared to the unprocessed condition. Preference rating was done under office-like room conditions only, with the three interfering noise sources (S0N90/180/270). Each pair of algorithms was presented at three different SNRs (0, +5, and +10 dB), and every presentation was conducted twice (test and retest). This gave a total of 30 ratings.

During the PR, the subject could listen to the algorithms as long as needed and could toggle between the algorithms as often as wanted. After indicating a preference, the subject had to rate how much better the preferred algorithm was compared to the other one. This rating can be interpreted as the confidence of the subjective preference judgments. The outcome of the test is the amount of preference for an algorithm over the unprocessed condition. This preference score can vary between "very much worse" (scored as −5) and "very much better" (scored as +5). The PR test used a forced-choice paradigm, so equal preference (that would be represented by score 0) was not an option.

The preference-rating data were rescaled with a linear Gaussian model (LGM) in order to obtain a perceptually correct scale. The data were evaluated according to the method of Dahlquist and Leijon (2003), where it is assumed that the perceptual quality of each algorithm has a specific neural representation in the listener's sensory system. This representation can be modeled as the outcome of a normally distributed (Gaussian) random variable X with mean μ_i, whenever algorithm A_i is presented, and a constant variance σ^2. The LGM places the algorithms on an interval scale.

It is assumed that each paired-comparison decision is determined by the outcome of two independent random variables (X_1, X_2) with Gaussian distributions with variance σ^2 and means (μ_i, μ_j), whenever algorithm pair (A_i, A_j) is presented. The LGM estimates the (unknown) parameters, leading to a nonlinear and multidimensional maximization problem, which can be solved using routines provided by the MATLAB optimization toolbox. The LGM provides the limits of the estimated confidence intervals. These limits can be used to evaluate the differences between two specific algorithms in terms of the five categories that were used by the subjects to rate the algorithms. As the unprocessed condition was used as a reference, all results were shifted so that the reference was placed at scale value zero. The LGM was applied to each set of data for every subject. Afterwards the results for the individuals were pooled.

F. Statistical analyses

Statistical analyses were carried out with the SPSS software. To test whether the distribution of the variables was normal, Kolmogorov–Smirnov tests were used. To analyze the SRT data, repeated-measures analyses of variance (ANOVAs) were carried out. For tests of within-subjects effects, lower-bound corrections for violations of sphericity were applied. *Post hoc* tests consisted of pairwise comparisons. To keep the type I error rate across all comparisons at 0.05, Bonferroni correction was applied. For LES and PR data, non-parametric statistics were applied. When several conditions were compared and consequently several tests were required, the significance level was also adjusted using Bonferroni correction (unless mentioned differently).

V. RESULTS

A. Test-retest reliability

Test and retest scores for SRT, LES, and PR were compared for results obtained in the office-like room, with three interfering noise sources. Test and retest scores for all algorithms were taken together. In this way, for SRT, only one paired-comparison was carried out (using a paired-samples

t-test). For LES and PR, data were compared for each tested SNR with a non-parametric Wilcoxon signed rank test.

Test-retest scores for SRT were significantly different ($p \leq 0.001$). The mean difference was 0.8 dB with the retest scores being lower than the test scores. This might be due to a learning effect. For LES there was a significant difference between test and retest at −10, −5, and 0 dB SNR ($p \leq 0.001$) with the retest scores being, respectively, 0.1, 0.2, and 0.3 points higher than the test scores. For PR, test and retest were not significantly different at any SNR ($p > 0.5$).

Test-retest reliability was assessed by calculating the variability of the scores, measured as the within-subjects standard deviation of the scores (σ_w) with the formula

$$\sigma_w = \sqrt{\frac{1}{2}\sum_{i=1}^{n}\frac{(x_{i1}-x_{i2})^2}{n}},$$

where x_{i1} is the ith test score, x_{i2} is the ith retest score, and n is the total number of scores to compare. For SRT, σ_w of the absolute scores was 1.8 dB. For LES, which is scored on a scale from 0 to 6, σ_w was 0.6, 0.9, 1.2, 1.0, and 0.7 points at −10, −5, 0, 5, and 10 dB SNR, respectively. The reliability was lower for the middle SNRs. At the lowest and the highest SNRs, subjects tended to give the maximum and minimum scores, respectively. At the extremes the variability between subjects, as well as within subjects, was thus small. For PR, which is scored on a scale from −5 to +5, σ_w was 1.9, 2.0, and 2.0 points for 0, +5, and +10 dB SNR, respectively.

For further analyses of the SRT results, test-retest was included in the ANOVAs as a within-subjects factor. For further analyses of LES and PR, test and retest scores were averaged prior to analysis.

B. Speech reception threshold tests

1. Office-like room and three interfering noise sources

To compare the results of the different subject groups and test sites, a mixed-model ANOVA was carried out including the absolute SRT scores measured in the office-like room with three noise sources, because these measurements were performed at all test sites. The analysis included two within-subjects factors (test-retest and algorithm) and two between-subjects factors (subject group and test site).

Both between-subjects factors were highly significant ($p \leq 0.001$). Pairwise comparisons of the test sites showed that the SRTs (mean ± SE) averaged over all subject groups and all algorithms for the German-speaking test sites DE (−4.7 ± 0.4 dB SNR) and CH (−3.4 ± 0.5 dB SNR) were not significantly different ($p = 0.290$). Similarly the average SRTs for the Dutch-speaking test sites BE (0.2 ± 0.4 dB SNR) and NL (−0.7 ± 0.4 dB SNR) were not significantly different ($p = 0.659$). Differences between German-speaking and Dutch-speaking test sites were highly significant ($p \leq 0.001$). As expected, pairwise comparisons of the subject groups (averaged over all test sites) showed that the NH subjects (−5.2 ± 0.4 dB SNR) had significantly lower SRTs than the HI subject groups ($p \leq 0.001$). The difference between the two HI subject groups was not significant ($p = 1.000$). The average SRT was −0.5 ± 0.4 dB SNR for the group with flat hearing losses and −0.8 ± 0.4 dB SNR for the group with sloping hearing losses.

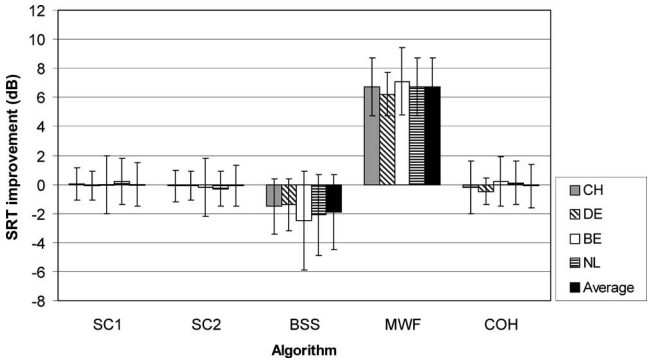


FIG. 1. Mean SRTs (and standard deviations) relative to those for the unprocessed condition in an office-like room with three noise sources. Results are shown for the different test sites averaged over all subject groups.

The main effect of test-retest ($p \leq 0.001$) was significant. The mean difference between test and retest data was 0.8 ± 0.1 dB with lower SRTs for the retest. There were no significant interaction effects between the factor test-retest and any other factors in this analysis ($p > 0.24$). The effect of test-retest is thus independent of algorithm, test site, and subject group.

There was a main effect of algorithm ($p \leq 0.001$), but the factor algorithm did not interact with other factors ($p > 0.23$). The differences between the algorithms are thus similar for test and retest scores, and for all subject groups and test sites. Pairwise comparisons of the SRTs for the five algorithms under test and the unprocessed condition showed that only BSS and MWF differed significantly from the unprocessed condition. The improvement with MWF was 6.6 ± 0.2 dB ($p \leq 0.001$). With BSS the SRT worsened by 1.9 ± 0.3 dB ($p \leq 0.001$).

In what follows, we focus on the SRT changes for the different algorithms relative to the SRT for the unprocessed condition, as the SRT changes or relative scores did not interact with subject group and test site. Figures 1 and 2 show the mean SRT improvements, averaged over subjects and test/retest, for the different test sites and subject groups, respectively. Positive numbers in the figures indicate an improvement.

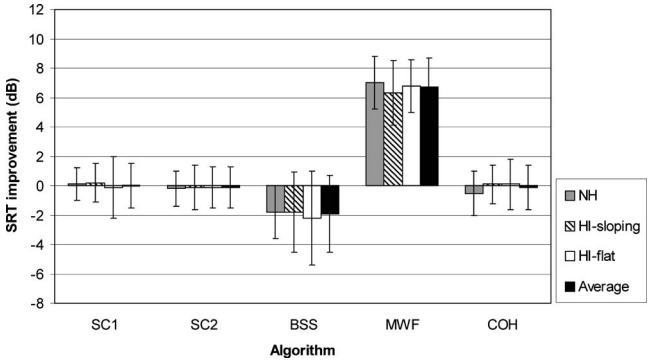


FIG. 2. As Fig. 1, but for the three subject groups averaged over all test sites.

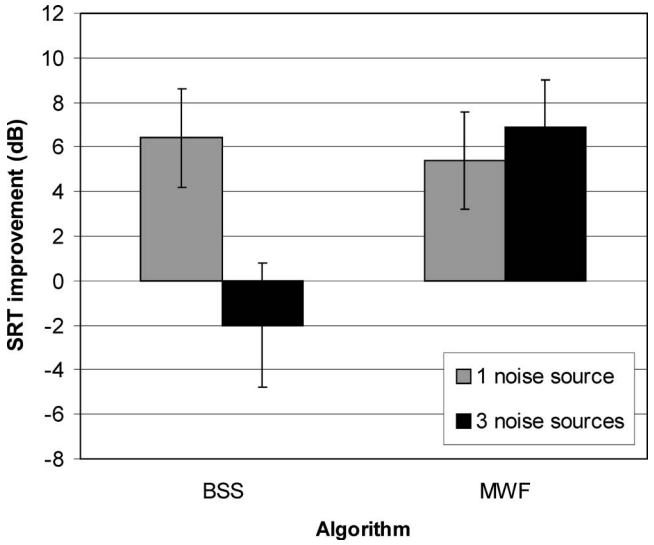


FIG. 3. Comparison of results for two noise scenarios. Mean SRTs (and standard deviations) relative to those for the unprocessed condition in an office-like room are shown. Results are averaged over three subject groups and three test sites.

2. Evaluation of BSS and MWF in two different noise scenarios

At three test sites (BE, NL, and CH), additional measurements were carried out in the office-like room with one noise source for the BSS and MWF. An ANOVA on the SRT improvements with three factors (test-retest, noise scenario, and algorithm) showed no effect of test-retest ($p=0.617$). There was a main effect of noise scenario and algorithm and an interaction between the two ($p \leq 0.001$). As shown in Fig. 3, the performance of the MWF improved significantly from 5.4 (SD=2.2 dB) to 6.9 dB (SD=2.1 dB) when adding more noise sources. The BSS performed well with one interfering noise source (6.4 ± 2.2 dB improvement), but led to worse performance for three uncorrelated noise sources (-2.0 ± 2.8 dB).

3. Comparison of test rooms

At two test sites (BE and DE), the algorithms were also evaluated in a highly reverberant room using three interfering noise sources. The SRTs in the highly reverberant room were on average 6.7 dB higher at BE and 5.2 dB higher at DE than the SRTs in an office-like room. A mixed-model ANOVA on the SRT improvements was carried out with the within-subjects factors test-retest (two levels), room (two levels), and algorithm (five levels). Subject group and test site were between-subjects factors.

The factors subject group and test site were not significant ($p > 0.05$). The main effect of room was not significant ($p=0.697$), but there was a significant interaction between room and algorithm ($p=0.009$). The origin of this interaction effect is shown in Fig. 4. The SRT improvement for the MWF was lower for the reverberant room (5.4 ± 2.2 dB) than for the office-like room (6.7 ± 2.0 dB), while the effects for the other algorithms were comparable across rooms.

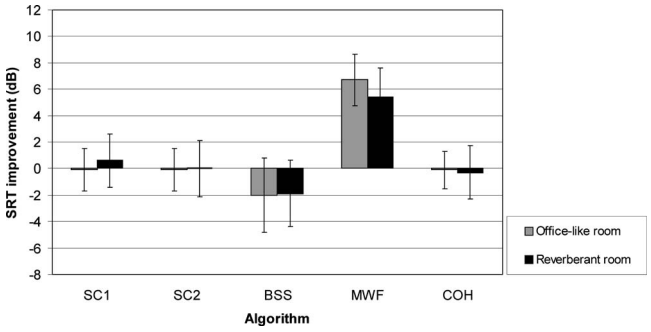


FIG. 4. Comparison of results for two types of rooms. Mean SRTs (and standard deviations) relative to those for the unprocessed condition are shown. Results are averaged over three subject groups and two test sites.

C. LES

Unlike the SRT scores, the LES scores were not normally distributed. Moreover, the distributions were very different for the different SNRs. Therefore, non-parametric tests were used.

1. LES for the unprocessed condition

The LES scores for the unprocessed condition (average of test and retest scores) were analyzed in order to assess whether there were baseline differences in LES scores between subject groups, test sites, and SNRs. The LES scores for the unprocessed condition are depicted in Fig. 5 as a function of SNR for the four test sites (left panel) and three subject groups (right panel). For each SNR, the effect of test site and subject group was evaluated with a non-parametric Kruskal–Wallis test. The effect of test site was significant at all SNRs ($p \leq 0.005$), except at -10 dB. In general, the effect was similar to the effect observed in the SRT scores. The SRTs for the Dutch-speaking subjects were higher than for the German-speaking subjects. Similarly, the Dutch-speaking subjects required more effort than the German-speaking subjects at the same SNR.

The effect of subject group was significant at -5, +5, and +10 dB SNR ($p \leq 0.005$). At these SNRs, Mann–Whitney tests were carried out to investigate the differences between groups. At all three SNRs the HI groups did not differ significantly ($p > 0.4$), but the NH group differed significantly from both HI groups ($p \leq 0.011$). Again, this effect is similar to the SRT scores for the unprocessed condition: HI subjects required a higher SNR to reach 50% intelligibility than NH subjects. Accordingly, at the same SNR, HI subjects required more effort than NH subjects. No difference was observed between the two HI groups for SRT or LES.

2. LES scores relative to the unprocessed condition

We focus on scores relative to those for the unprocessed condition. A positive relative score means that less effort was required than for the unprocessed condition. The effect of subject group and test site on the relative LES scores was assessed with Kruskal–Wallis tests for each algorithm at each SNR. After Bonferroni correction for multiple testing, only one variable (the MWF at -10 dB SNR) showed an effect of

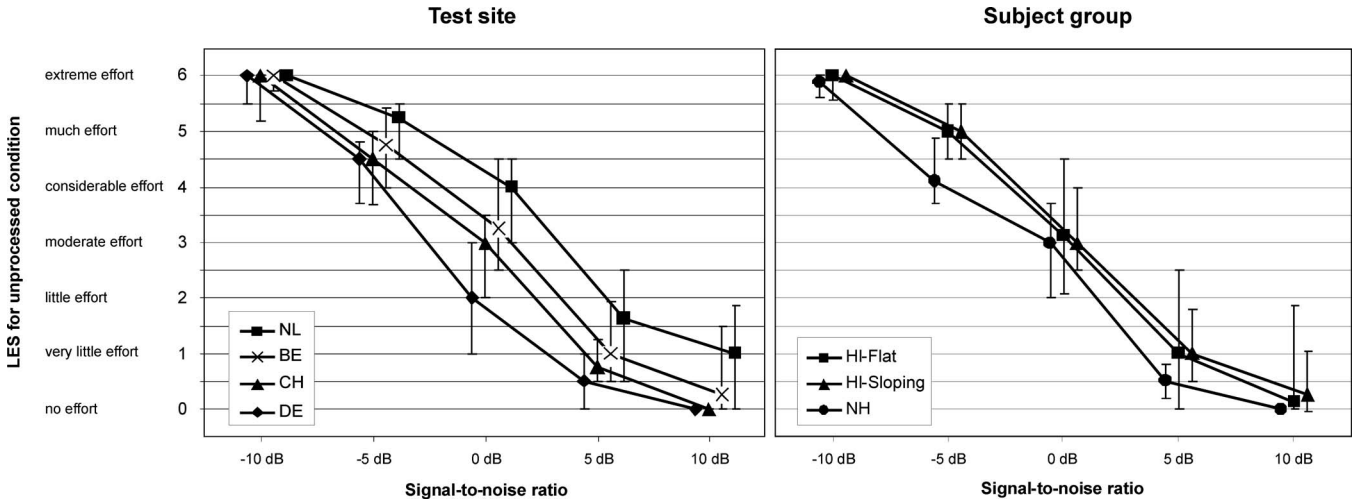


FIG. 5. Median LES for the unprocessed condition for different SNRs. The effects of test site and subject group are shown in the left and right panels, respectively. Error bars represent first and third quartiles.

subject group. The effect of test site, however, was significant for many variables. Therefore, Fig. 6 shows the LES difference scores for each test site, but averaged over subject groups. Because the results were not normally distributed, medians and quartiles are shown. Results are only shown for 0 dB SNR as the largest effects were obtained at this SNR. To assess whether LES scores for the algorithms were different from those for the unprocessed condition, Wilcoxon signed ranks tests were carried out. As 25 comparisons were needed, the significance level was decreased to 0.002 ($p = 0.05/25$). Algorithm scores that were significantly different from those for the unprocessed condition are indicated with an asterisk. The absolute effort scores for the algorithms can be estimated by adding the relative scores to the average effort scores for the unprocessed condition as shown in Fig. 5. As becomes clear from Fig. 6, results differed across test sites. However, general trends were similar. Figure 7 illustrates the effect of SNR on the relative LES scores, averaged over subject group and test site. Overall, the MWF required less effort than the unprocessed condition. This effect was largest at low SNRs, i.e., in the more difficult conditions. BSS was the only algorithm that required significantly more efforts than the unprocessed condition, and this was the case at SNRs from −5 to +10 dB SNR. This effect, however, did not occur at all test sites. The other algorithms did not have a large effect on the effort scores. Only at 0 dB SNR did SC1, SC2, and COH have an overall positive effect on the effort scores.

D. PRs

1. Win count

The PRs were initially analyzed by counting the percentage of wins, which is the percentage of times a given algorithm was preferred over the unprocessed condition. If subjects had no preference, the percentage of wins would be 50%. Figure 8 shows the percentage of wins for the five algorithms for the three tested SNRs. SC1, SC2, MWF, and COH were significantly preferred over the unprocessed condition. For BSS, subjects preferred the unprocessed condition (binomial test, $p \leq 0.001$). The general trends were the same for the three SNRs.

To compare the percentage of wins between the different algorithms, McNemars χ^2 test was applied. To limit the num-

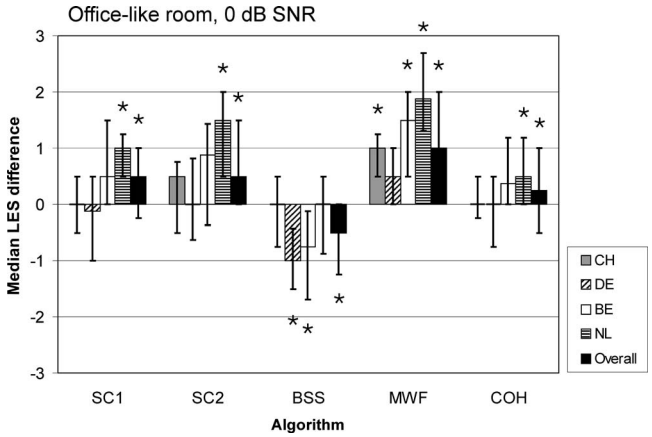


FIG. 6. LES relative to the LES for the unprocessed condition for 0 dB SNR. The error bars show the first and third quartiles. Algorithm scores that are significantly different from those for the unprocessed condition are indicated with an asterisk.

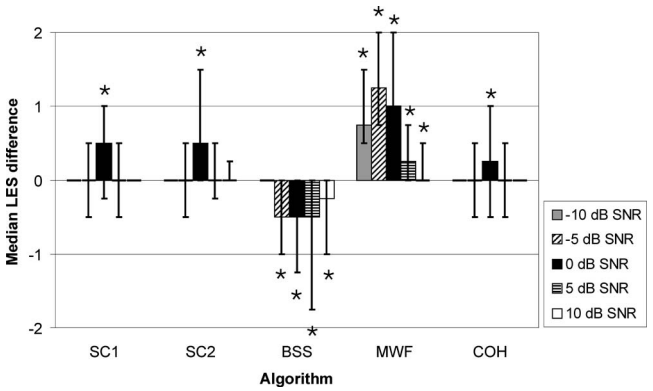


FIG. 7. LES relative to the LES for the unprocessed condition for SNRs from −10 to +10 dB, averaged over all subjects. The error bars show the first and third quartiles. Algorithm scores that are significantly different from those for the unprocessed condition are indicated with an asterisk.

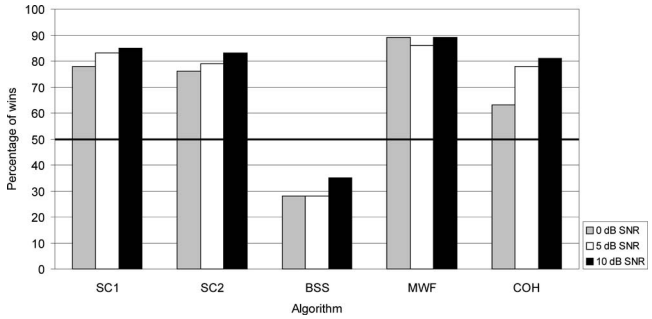


FIG. 8. PR results expressed as the percentage of wins for the algorithms compared to the unprocessed condition at three SNRs.

ber of comparisons, the total of wins across SNRs was used. Except for the differences between SC1 and SC2, and SC2 and COH, all differences were significant ($p \leq 0.001$, with a significance level of $p=0.005$ after Bonferroni correction). The MWF had the highest percentage of wins. However, the difference between MWF and the other algorithms was very small.

2. Degree of preference

The PR scores were not normally distributed, and thus non-parametric statistics were applied. The PR data were rescaled with a LGM. A Kruskal–Wallis test on the rescaled data of each algorithm showed no significant effect of test site ($p > 0.4$). A significant effect of subject group was found only for COH ($p=0.001$). COH was preferred more often than the unprocessed condition by the HI subjects, but not by the NH subjects. Figure 9 shows the effect of algorithm and SNR, averaged over the test sites and subject groups. For the noise scenario with three noise sources, MWF led to the highest preference score. SC1, SC2, and COH were also preferred over the unprocessed condition. BSS had a negative preference score, meaning that the unprocessed condition was more preferred. For each of the algorithms, the effect of SNR was significant (Friedman's ANOVA, $p \leq 0.001$). The degree of preference (positive or negative) was in general lower at 0 dB SNR than at the higher SNRs.

E. Correlations SRT-LES-PR

SRT, LES, and PR scores were obtained for all subjects and in the same test conditions. The correlation between results of the three tests was determined.

1. Relative scores

For PR, only scores relative to the unprocessed condition were available. Consequently, correlations between PR, SRT, and LES scores were calculated for the relative scores. As not all variables were normally distributed, Spearman correlations were calculated. Doing this for each algorithm separately did not result in any significant correlation. This is not surprising, as the variability between subjects (and thus the data range) was smaller for the relative scores than for the absolute scores. Several algorithms did not show any improvement for SRT and LES scores. This restricted range of the variables affects the correlations.

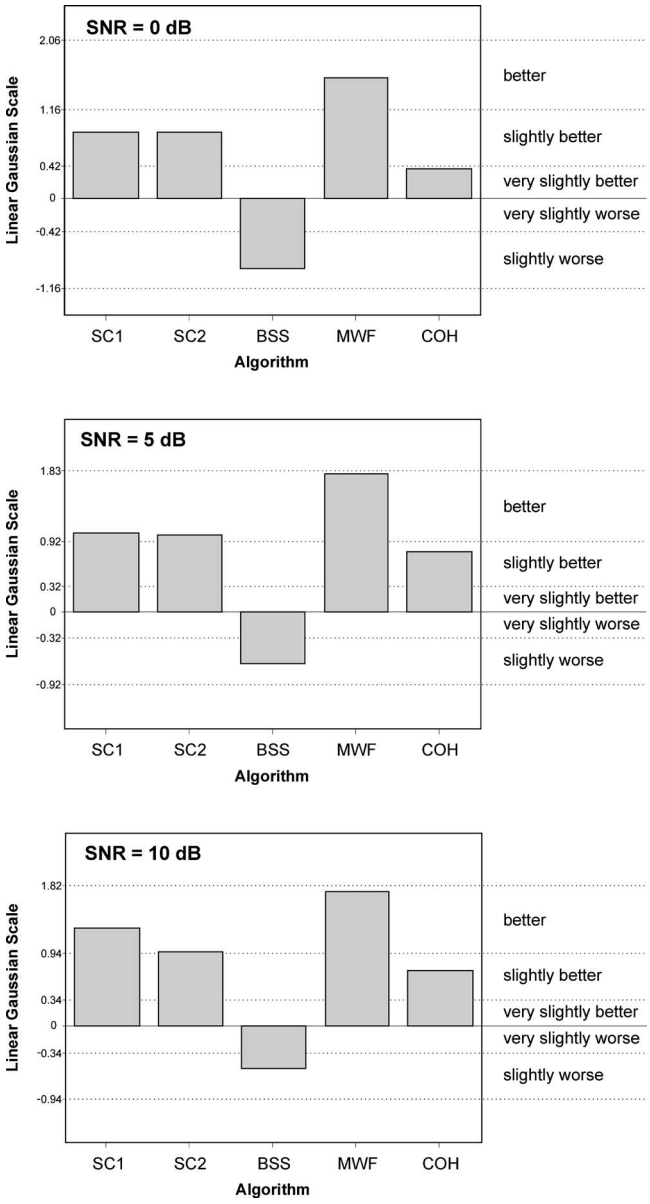


FIG. 9. PR scores for the different SNRs averaged over the test sites and subject groups. The medians (for the rating of the algorithms and the estimated confidence intervals) of the individual (subject) data, transformed with a LGM, are plotted with bars. The offset along the LGM scale is adjusted to zero for the unprocessed condition for every subject individually. Additionally, the categories are indicated.

Correlations were also calculated for all algorithms together, to investigate if subjects rated algorithms in a consistent way over the different tests. The relative SRT scores were significantly correlated ($p \leq 0.001$) with relative LES and PR scores at all tested SNRs. For LES the correlations ranged from 0.192 at 10 dB SNR to 0.418 at −5 dB SNR. For PR the correlations ranged from 0.316 at 10 dB SNR to 0.379 at 0 dB SNR. These results indicate that improvement in speech intelligibility is weakly associated with improvement in listening effort and degree of preference. The average SRT over all algorithms was between 0 and −5 dB SNR for the different groups, so it is not surprising that the highest correlations were obtained at these (low) SNRs. PR and LES relative scores were also significantly correlated ($p \leq 0.001$),

although correlations were low. The Spearman correlations were 0.237, 0.179, and 0.200 at 0, 5, and 10 dB SNR, respectively.

2. Absolute scores

The absolute LES and SRT scores were also compared. Overall correlations (for all algorithms and the unprocessed condition together) ranged from 0.439 to 0.547 ($p \leq 0.001$). The highest correlation was found at −5 dB SNR. Correlations for each algorithm separately were all significant at $p \leq 0.05$ (without correction for multiple comparisons). At −5 dB SNR, Spearman correlations were 0.454, 0.348, 0.442, 0.499, 0.277, and 0.437 for the unprocessed, SC1, SC2, BSS, MWF, and COH conditions, respectively ($p \leq 0.004$). These correlations indicate that lower SRTs (better speech intelligibility) are associated with lower LES scores (less listening effort).

VI. DISCUSSION

A. Effect of test site

One goal of this study was to evaluate how comparable the results were for the different test sites. To ensure comparability of the results across the four test sites, special attention had to be paid to the test protocol. On the one hand, the protocol had to be flexible to allow for inherent variations, e.g., of the specification of the test rooms and the auditory profile of the subjects. On the other hand, the protocol needed to be sufficiently strict to minimize the impact of unknown factors on the test results.

For the absolute SRT scores, a significant effect of test site was observed. The overall SRT in multitalker babble noise for the German-speaking test sites was about 3.6 dB lower than the SRT for the Dutch-speaking test sites. This is not surprising, as each speech material has its own reference psychometric curve with an average SRT and slope. The SRT for the German OLSA sentences in stationary speech-weighted noise for NH subjects is lower than the SRT in the same conditions for the Dutch VU-sentences. This absolute shift in SRT scores can be taken into account by considering the improvements relative to the SRTs for the unprocessed condition. In this study, care was taken to use a uniform procedure for, e.g., calibration of the setup and fitting of the hearing aids. A common evaluation platform was used, test procedures were thoroughly discussed and evaluated through a series of pilot tests, and test instructions were uniform across test sites. Other parameters, such as the specific type of loudspeaker, sound card, speech material, and exact room characteristics such as size and reverberation time, differed between test sites. Variations in these parameters (controlled to a certain extent) did not significantly influence the effects of the algorithms on SRTs.

For PR, there were no significant differences between the test sites. For LES, in contrast, there were differences across the test sites. For the unprocessed condition, the pattern of the LES scores resembled the pattern of the corresponding SRT scores. Lower absolute SRTs corresponded with lower effort scores. For LES scores, the differences between test sites were not completely eliminated by taking the relative scores. General trends, however, were the same across test sites. It was mainly the size of the effects that differed. It is not clear what the cause of these differences was.

B. Effect of subject group

The absolute SRT and LES scores showed differences between subject groups. As expected, HI subjects had higher (worse) absolute SRT and LES scores than NH subjects. Nevertheless, the relative SRT and LES scores were not significantly different between subject groups. For PR, results were also very similar between subject groups. It is reasonable to assume that the noise suppression algorithms succeed in improving the SNR by some decibels, at least for positive SNRs. As they operate on the noisy speech signals before they are delivered to the NH or HI subject, this could result in similar perceptual improvements for all subject groups.

Besides the perceptual evaluation of the algorithms, a physical evaluation was performed (not reported in this paper), where the algorithm effects were evaluated for several different "auditory profiles," i.e., types of hearing impairment. The algorithm performance was quantified by objective performance measures that use established auditory models to account for one or more of the following fundamental effects of hearing impairment: loss of audibility at low input levels, loudness recruitment, reduced frequency resolution, and reduced ability to extract supra-threshold speech cues for speech recognition. The performance measures included (among others) the SII-improvement, which is a standardized and commonly used performance measure for predicting speech intelligibility in non-fluctuating noise (ANSI S3.5-1997, 1997) and the segmental SII-improvement, a modified version where the SII is first calculated for each short-time segment, and these values are then averaged over the duration of the test signal. According to this physical evaluation, the algorithms did show differences in performance for different auditory profiles (Eneman *et al.*, 2008a, 2008b). NH subjects were expected to benefit more from the algorithms than HI subjects. Additionally, a lower benefit was expected for the flat than for the sloping audiogram configurations. The perceptual results, which show no differences between NH and HI listeners, appear inconsistent with the physical evaluation, but there are several possible explanations for this discrepancy. First, in the perceptual study, the groups were defined by audiogram information only. The HI subjects were assigned to the HI-S or HI-F group on the basis of slope of the audiogram between 500 and 4000 Hz, but they might have differed in other ways. Other aspects of the auditory profile might affect the perceptual scores. Second, the SRT (defined as the SNR for 50% intelligibility) and SRT improvement were determined at very different SNRs for the NH and HI subjects. It was postulated that the algorithms would perform worse in terms of SRT improvement at low (negative) SNRs, e.g., the speech-noise detector of MWF would perform worse at low SNRs. The algorithms may decrease the noise level at these low SNRs, but this may be accompanied by a decrease in the speech level as well. Indeed, Eneman *et al.* (2008a, 2008b)

showed that the segmental SII-improvement of the MWF worsened with decreasing SNR below 0 dB. This might explain the decreased algorithm benefit measured for NH subjects compared to the benefit measured for the HI groups, whose SRTs were higher. To conclude, the discrepancy between physical and perceptual results emphasizes once more the importance of perceptual evaluations using NH and HI subjects.

C. Algorithm performance

1. SRT results

In the pseudo-diffuse noise scenario, MWF was the only algorithm that provided a significant SRT improvement relative to the unprocessed condition. Averaged over test sites and subject groups, an improvement of 6.6 dB was obtained with the MWF. The single-channel noise reduction methods, SC1 and SC2, did not significantly affect SRTs. This is noteworthy, since in single-channel noise reduction the competing signal cannot be removed without distorting the desired signal. Single-channel noise reduction is a challenging task, in particular, at the low SNRs at which the SRT is measured. The BSS had a negative effect and increased SRTs relative to those for the unprocessed condition, by on average 1.9 dB. This is not really surprising, as the two-channel BSS algorithm used was designed to separate two distinct point sources only. At three test sites, BSS and MWF were also evaluated using a single point-source masker. The multimicrophone enhancement techniques BSS and MWF were able to take advantage of the spatial diversity of the setup and to improve speech intelligibility in this noise scenario by, on average, 6.4 and 5.4 dB, respectively. Whereas the MWF algorithm performed well in both noise scenarios (with one and three noise sources), the BSS algorithm appears suited for the single point-source masker only.

The additional measurements in the reverberant room showed that the SRT improvements were very similar for the two rooms. Only for MWF was the relative SRT score significantly different for the two rooms. In the reverberant room, the improvement in SRT reduced by 1.3 dB. Nevertheless, even in these very demanding listening circumstances, an improvement of 5.4 dB was obtained.

2. Comparison of SRT, LES, and PR

As a consequence of the adaptive procedure, SRTs and SRT improvements were measured at different SNRs depending on the subjects' hearing status. Additionally, for many subjects, the SRT was at a negative SNR. This is a disadvantage of the speech intelligibility measurements, as it is known that several signal enhancement algorithms perform differently at low and high SNRs. Nonetheless, speech intelligibility improvements are most required in difficult listening situations (e.g., at 50% intelligibility), which occur at different SNRs depending on the hearing loss. Therefore, SRT measurements are considered to be highly relevant in the perceptual evaluation of speech enhancement algorithms. In addition to the SRT measurements, LES and PR can be used to evaluate the algorithm improvements at high input SNRs, as the measurements can be obtained for SNRs up to +10 dB.

For the relative LES scores, there were differences between test sites. However, some general observations could be made. In the pseudo-diffuse noise scenario, the MWF required less effort than the unprocessed condition, especially at the lower SNRs. At the higher SNRs, speech was easy to understand in the unprocessed condition, and thus the MWF did not decrease the listening effort much. At some test sites, the BSS led to a negative effect on listening effort, mainly at the higher SNRs, although variability was large. This was not unexpected since the LES scores (and PR scores) were obtained in a scenario with three uncorrelated noise sources, whereas the two-channel BSS algorithm was designed to separate two distinct point sources.

The LES revealed improvements for algorithms, which did not give improved SRTs, analogous to the results of Marzinzik and Kollmeier (1999) and Walden *et al.* (2000). These researchers showed a significant positive effect of noise reduction algorithms on LES, although no improvement in speech intelligibility was shown by SRT tests. Similarly, in the current study, an improvement in listening effort for SC1, SC2, and COH was found at 0 dB SNR, although no SRT improvement was measured for these algorithms. At the other SNRs, LES did not reveal improved listening effort for these three algorithms.

For the PR results, preference was higher for MWF, SC1, SC2, and COH than for the unprocessed condition, at all three SNRs. So, PR did reveal differences for a number of algorithms. The preference for the algorithms was higher at the high SNRs (+5 and +10 dB) than at 0 dB. The increasing PR scores with increasing SNR for SC1, SC2, and COH are most probably due to less (audible) speech and noise distortion at higher SNRs. Analogous results were obtained in a study of Ricketts and Hornsby (2005). They used a similar paired-comparison approach to determine preference. Although the single-channel noise reduction in a commercial hearing aid did not affect speech intelligibility, there was a strong preference for the noise reduction algorithm using both low-level and high-level noise. Natarajan *et al.* (2005) also reported increased speech quality ratings, while no measurable changes in intelligibility occurred. A study of Alcántara *et al.* (2003), however, showed no preference for a commercial noise reduction algorithm.

The discrepancy between LES and PR has several possible causes. First, different dimensions were rated. LES is based on a rating of the effort in understanding speech in a noisy situation and PR gives an indication of the preference for an algorithm in a noisy situation, which is much broader and includes the general quality of sounds processed using an algorithm. Second, the psychophysical methods used for LES and PR are different. LES makes use of a categorical magnitude rating, whereas PR is based on paired-comparison. In addition to this methodological difference (and as a consequence), the LES scores show clear floor and ceiling effects at the lowest and highest SNRs. Because of these floor and ceiling effects it might be more difficult to

measure differences between the algorithms and the unprocessed condition. This is avoided in the paired-comparison procedure used for the PR test.

VII. CONCLUSIONS

In this multicenter study, five different signal enhancement approaches were evaluated through listening tests with large numbers of NH listeners and hearing aid users with flat and sloping moderate hearing losses. The SRT improvements and PR scores for the different signal processing strategies were comparable across the test sites, despite differences in test material. For LES, slightly different results were obtained across the test sites. Additional investigations are necessary to interpret these differences. For the SRT improvements, LES improvements, and PR scores, outcomes were very similar for the different subject groups. In a pseudo-diffuse noise scenario, MWF was the only algorithm that provided a significant SRT improvement. The BSS algorithm seemed suited only for the single point-source scenario. Unlike the lack of improvement in SRT for SC1, SC2, and COH, the PR did show preference for these algorithms over the unprocessed condition at all tested SNRs and the LES showed decreased listening effort at 0 dB SNR for the same algorithms. This study shows that perceptual evaluations of signal enhancement algorithms can be obtained in a precise, robust, reproducible, and comparable manner across different research groups and subject groups when an appropriate test protocol is used.

ACKNOWLEDGMENTS

The authors thank Sofie Jansen, Niklas Harlander, Elja Schaaf, Andrea Kegel, Immo Ress, Melanie Zokoll, Müge Kaya, and Jessica Beeken for their help in collecting the data. Ann Spriet is a postdoctoral researcher funded by the F.W.O.-Vlaanderen. The authors are also grateful to the reviewers and to Brian Moore for their helpful comments. This work was supported by a grant from the European Union FP6, Project No. 004171 HEARCOM.

Aichner, R., Buchner, H., Yan, F., and Kellermann, W. (**2006**). "A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments," Signal Process. **86**, 1260–1277.

Alcántara, J. I., Moore, B. C. J., Kühnel, V., and Launer, S. (**2003**). "Evaluation of the noise reduction system in a commercial digital hearing aid," Int. J. Audiol. **42**, 34–42.

ANSI S3.5-1997 (**1997**). "American National Standard: Methods for calculation of the speech intelligibility index," American National Standards Institute, New York.

Bentler, R., Wu, Y.-H., Kettel, J., and Hurtig, R. (**2008**). "Digital noise reduction: Outcomes from laboratory and field studies," Int. J. Audiol. **47**, 447–460.

Blamey, P. J. (**2005**). "Adaptive dynamic range optimization (ADRO): A digital amplification strategy for hearing aids and cochlear implants," Trends Amplif. **9**, 77–98.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Buchner, H., Aichner, R., and Kellermann, W. (**2004**). "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication*, edited by Y. Huang and J. Benesty (Kluwer Academic, Norwell, MA), pp. 255–293.

Buchner, H., Aichner, R., and Kellermann, W. (**2005a**). "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," IEEE Trans. Speech Audio Process. **13**, 120–134.

Buchner, H., Aichner, R., Stenglein, J., Teutsch, H., and Kellermann, W. (**2005b**). "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Philadelphia, PA.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in *The Vanderbilt Hearing Aid Report II*, edited by G. Studebaker, F. Bess, and L. Beck (York, Parkton, NC), pp. 295–300.

Dahlquist, M., and Leijon, A. (**2003**). "Paired-comparison rating of sound quality using MAP parameter estimation for data analysis," in Proceedings of the Tutorial and Research Workshop on Auditory Quality of Systems (AQS), Mont Cenis, Germany, pp. 79–84.

Dillon, H. (**2001**). *Hearing Aids* (Boomerang, Sydney, Australia).

Doclo, S., Spriet, A., Moonen, M., and Wouters, J. (**2007**). "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," Speech Commun. **49**, 636–656.

Doclo, S., Spriet, A., Wouters, J., and Moonen, M. (**2005**). "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, edited by J. Benesty, S. Makino, and J. Chen, (Springer, Berlin), pp. 199–228.

Eneman, K., Leijon, A., Doclo, S., Spriet, A., Moonen, M., and Wouters, J. (**2008a**). "Auditory-profile-based physical evaluation of multimicrophone noise reduction techniques in hearing instruments," in *Advances in Digital Speech Transmission*, edited by R. Martin, U. Heute, and C. Antweiler, (Wiley, Hoboken, NJ), pp. 431–458.

Eneman, K., Luts, H., Wouters, J., Büchler, M., Dillier, N., Dreschler, W., Froehlich, M., Grimm, G., Hohmann, V., Houden, R., Leijon, A., Lombard, A., Mauler, D., Moonen, M., Puder, H., Schulte, M., Spriet, A., and Vormann, M. (**2008b**). "Evaluation of signal enhancement algorithms for hearing instruments," in Proceedings of the European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland, pp. 1–5.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121.

Francart, T., van Wieringen, A., and Wouters, J. (**2008**). "APEX 3: A multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods **172**, 283–293.

Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (**1993**). "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," J. Acoust. Soc. Am. **94**, 3009–3010.

Grimm, G., Guilmin, G., Poppen, F., Vlaming, M. S. M. G., and Hohmann, V. (**2009a**). "The personal hearing system—A software hearing aid for a personal communication system," EURASIP J. Adv. Signal Process **2009**, 591921.

Grimm, G., Herzke, T., Berg, D., and Hohmann, V. (**2006**). "The master hearing aid: A PC-based platform for algorithm development and evaluation," Acta Acust. Acust. **92**, 618–628.

Grimm, G., Hohmann, V., and Kollmeier, B. (**2009b**). "Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme," IEEE Trans. Audio, Speech, Lang. Process. **17**, 1408–1419.

Hu, Y., and Loizou, P. C. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**, 498–509.

Kochkin, S. (**2005**). "MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age," Hear. J. **58**, 30–43.

Martin, R. (**2001**). "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process. **9**, 504–512.

Marzinzik, M., and Kollmeier, B. (**1999**). "Development and evaluation of single-microphone noise reduction algorithms for digital hearing aids," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier, (World Scientific, Singapore), pp. 279–282.

Mauler, D., and Martin, R. (**2006**). "Noise power spectral density estimation on highly correlated data," in Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, pp. 1–4.

Natarajan, A., Hansen, J. H. L., Arehart, K. H., and Rossi-Katz, J. (**2005**). "An auditory-masking-threshold-based noise suppression algorithm GMMSE-AMT[ERB] for listeners with sensorineural hearing loss," EURASIP J. Appl. Signal Process. **2005**, 2938–2953.

Noble, W., and Gatehouse, S. (**2006**). "Effects of bilateral versus unilateral

hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ)," Int. J. Audiol. **45**, 172–181.
Ricketts, T. A., and Hornsby, B. W. Y. (**2005**). "Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction," J. Am. Acad. Audiol. **16**, 270–277.
Spriet, A., Moonen, M., and Wouters, J. (**2004**). "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," Signal Process. **84**, 2367–2387.
Spriet, A., Moonen, M., and Wouters, J. (**2005**). "Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids," IEEE Trans. Signal Process. **53**, 911–925.
Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., van Wieringen, A., and Wouters, J. (**2007**). "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM(TM) in the Nucleus Freedom(TM) Cochlear Implant System," Ear Hear. **28**, 62–72.
Stone, M. A., and Moore, B. C. J. (**1999**). "Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses," Ear Hear. **20**, 182–192.
Vary, P., and Martin, R. (**2006**). *Digital Speech Transmission, Enhancement, Coding and Error Concealment* (Wiley, Chichester).
Versfeld, N., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Extension of sentence materials for the measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.
Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und evaluation eines satztests für die Deutsche sprache. Teil III: Evaluation des Oldenburger satztests (Development and evaluation of a sentence test for the German language. Part 3: Evaluation of the Oldenburg sentence test)," Z. Audiol **38**, 86–95.
Walden, B., Surr, R., Cord, M., Edwards, B., and Olson, L. (**2000**). "Comparison of benefits provided by different hearing aid technologies," J. Am. Acad. Audiol. **11**, 540–560.
Wittkop, T., and Hohmann, V. (**2003**). "Strategy-selective noise reduction for binaural digital hearing aids," Speech Commun. **39**, 111–138.

Speech production: Papers from a 2008 symposium

The five papers that follow are written versions of papers presented at the Sixth International Conference on Voice Physiology and Biomechanics (ICVPB), held in Tampere, Finland, on 6-9 August 2008. This series of conferences grew out of an earlier set of conferences on Vocal Fold Physiology, these having been held biannually since 1980. (The first of the present series took place in Evanston, Illinois, in May 1997.) The conferences have no proceedings as such, but the organizers endeavor to have authors of papers falling within distinct topical areas publish their work as a group. Four papers from the first conference were published as a group in the January 1999 issue of the Journal (Vol. 105:1). Six papers from the second conference were published as a group in the October 2000 issue (Vol. 108:4). The authors of the present five papers wanted to have them published in the Journal. Each of the papers was reviewed independently and subjected to the same rigorous peer review procedure as are all archival research articles submitted to the Journal. Four associate editors, David Berry, Avraham Hirschberg, Anders Lofqvist, and Douglas O'Shaughnessy handled the peer review processing.

Anders Lofqvist

Pharyngeal articulation in the production of voiced and voiceless fricatives[a)]

Michael I. Proctor,[b)] Christine H. Shadle, and Khalil Iskarous
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511

(Received 18 December 2008; revised 10 April 2009; accepted 19 April 2009)

A structural magnetic resonance imaging study has revealed that pharyngeal articulation varies considerably with voicing during the production of English fricatives. In a study of four speakers of American English, pharyngeal volume was generally found to be greater during the production of sustained voiced fricatives, compared to voiceless equivalents. Though pharyngeal expansion is expected for voiced stops, it is more surprising for voiced fricatives. For three speakers, all four voiced oral fricatives were produced with a larger pharynx than that used during the production of the voiceless fricative at the same place of articulation. For one speaker, pharyngeal volume during the production of voiceless labial fricatives was found to be greater, and sibilant pharyngeal volume varied with vocalic context as well as voicing. Pharyngeal expansion was primarily achieved through forward displacement of the anterior and lateral walls of the upper pharynx, but some displacement of the rear pharyngeal wall was also observed. These results suggest that the production of voiced fricatives involves the complex interaction of articulatory constraints from three separate goals: the formation of the appropriate oral constriction, the control of airflow through the constriction so as to achieve frication, and the maintenance of glottal oscillation by attending to transglottal pressure. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3299199]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt, 43.70.Gr [DAB] Pages: 1507–1518

I. INTRODUCTION

It has long been observed that voiced consonants are characterized by supralaryngeal configurations that differ from those used during the production of voiceless equivalents. Westbury (1983), for example, found that voiced stops "always exhibit active volume increases during their closures," but concluded that the precise nature of any such increase is not especially important, since the means by which the pharynx was expanded was found to vary widely. A cinefluorographic study of English stop cognates by Kent and Moll (1969) found that voiced stops were produced with larger supraglottal volumes caused by lengthening and expansion of the oropharynx, which was attributed to "the result of muscular action rather than passive responses of the vocal tract."

Many other studies have addressed pharyngeal behavior in voiced stops (e.g., Ladefoged, 1963; Rothenberg, 1968), but it has not been established which parts of the pharynx are responsible for voiced expansion, and there has been little treatment in the literature of the mechanisms of voicing in the production of fricatives.

At a first level of analysis, the voicing of fricatives is subject to the same physical constraints as any other voiced obstruent: for glottal excitation to occur, there must be sufficient airflow through the glottis. This is initiated by creating a pulmonically driven pressure differential across the glottis, but in order to *sustain* airflow, a mechanism for the reduction of supraglottal air pressure must be created. In the case of stop consonants, three mechanisms can be employed (Rothenberg, 1968, p. 91):

(i) a passive, pressure-actuated expansion of the walls of the supraglottal cavity,
(ii) a muscularly activated enlargement of the supraglottal cavity, and
(iii) (nasal) airflow through an incomplete (velopharyngeal) closure.

The production of fricatives typically involves no nasal airflow but is characterized by the continuous release of air from the supraglottal cavity through an oral constriction, so the same three mechanisms are available for absorption of glottal airflow. Another strategy might be to increase the aperture of the constriction during the production of voiced fricatives. However, if the speaker controls a similar rate of airflow through the fricative constriction for both members of a voiced-voiceless fricative pair, the question to be considered here is the extent to which the first two mechanisms come into play.

In the case of labial fricatives, supraglottal volume can be manipulated by adjusting the lips and cheeks. However, the more posterior the place of constriction, the fewer the options available for manipulating supraglottal volume. Considering also that the production of English fricatives requires precise articulatory control of the lips, tongue tip, and tongue body (Hardcastle, 1976), then it would appear that the pharynx—which is less directly involved in the generation of noise at the fricative constriction than the active articulators—should bear a greater functional load in the achievement of voicing requirements.

[a)]This paper is an expanded version of a talk presented by the second author at the Sixth ICVPB, Tampere, Finland, 6–9 August 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: proctor@haskins.yale.edu

The central question addressed in this paper is the way in which the pharynx is articulated during voicing; specifically, in the production of a voiced fricative:

- to what extent is pharyngeal volume actively manipulated or allowed to change passively?
- to what extent do competing constraints on the production of fricatives interact with the control of pharyngeal volume for voicing?

A. Supraglottal volume control during voicing

Rothenberg (1968) addressed in detail the ways in which "glottal airflow in a voiced plosive is often absorbed by a muscularly activated enlargement of the supraglottal cavity," arguing that there are three components of pharyngeal volume change, attributable to movements of the (i) anterior, (ii) vertical, and (iii) posterior-lateral boundaries of the pharynx.

Perkell (1965) reported expansion of the pharynx through forward movement of the base of the tongue by up to 50 mm, which Rothenberg (1968) estimated as contributing up to 6 ml of volume. Stetson (1951) described how vertical motion of the larynx-hyoid unit can elongate the pharynx in that dimension, and Perkell (1965) measured net vertical pharyngeal augmentation of the order of 2 mm during the production of voiced plosives. Based on these data, Rothenberg (1968) estimated that vertical displacement might be responsible for pharyngeal volume increases of up to 2 ml. The third mechanism of pharyngeal expansion, attributable largely to the action of the stylopharyngeous muscles, is estimated to contribute an additional 2 ml of volume, although Rothenberg (1968) did not base this figure on any phonetic data. In total, Rothenberg (1968) estimated that the combined action of these three components could expand the pharynx by up to 10 ml during the production of a voiced consonant, a factor of 20% for a pharyngeal volume of 50 ml.

Bell-Berti (1975) examined the activity of four muscle groups responsible for control of pharyngeal cavity size during the production of stops—levator palatini, superior constrictor, middle constrictor and sternohyoid—and found that all three subjects used at least one of the active expansion mode muscles significantly. She concluded that each subject used "a different arrangement of muscle activities to achieve the pharyngeal cavity expansion necessary for the continuation of glottal pulsing during voiced stop consonant occlusion." On the other hand, Magen *et al.* (2003) used magnetic resonance imaging (MRI) and x-ray cineradiography to examine the behavior of the posterior pharyngeal wall in Japanese and English speakers' vowel production, and found very little movement. The authors concluded that "the position of the posterior pharyngeal wall in this region can be eliminated as a variable, and the anterior portion of the pharynx alone can be used to estimate vocal cavities." The study did not include voiced-voiceless consonantal contrasts.

B. Supraglottal volume control in the production of fricatives

Although many of these same mechanisms are available to a speaker during the production of a voiced fricative, the task differs considerably from that of stop production because of the aerodynamic continuancy involved in producing frication. Ohala (1983) proposed that the difficulty of managing the competing aerodynamic requirements of frication and voicing might account for the comparative typological scarcity of voiced fricatives in comparison to the more commonly occurring voiced stops.

There are little phonetic data currently available with which to test such hypotheses and gain further insights into the details of voicing mechanisms in fricatives. In a cineradiographic study of English sibilants, Subtelny *et al.* (1972) found that "pharyngeal dimensions for /s/… were smaller than for /z/" within some vowel contexts, but these fricative pairs were not elicited in identical phonological contexts. Cineradiographic data collected by Perkell (1969) reveal pharyngeal augmentation during the production of /z/ due to advancement of the tongue dorsum 2 mm further than for /s/. A limitation of these studies is that the data are restricted to the midsagittal plane, and cannot be used to calculate pharyngeal volumes, nor to examine articulation in other dimensions.

In an MRI study of English fricatives, Narayanan *et al.* (1995) observed that "the tongue root tended to be more advanced in the case of the voiced fricatives when compared to their unvoiced counterparts …Tongue root advancement resulted in greater areas in the mid- and lower-pharyngeal regions (and) also influenced the epiglottic-vallecular volume to some extent." All fricatives in their study were elicited in a [ə_] context; however, it has been found that schwa is an inconsistently articulated context vowel, and that adjacent vowels can have a large influence on the position of the tongue when producing fricatives (Shadle *et al.*, 2008). This vocalic coarticulation effect is an important factor which must be controlled when considering the pharyngeal variation that can be attributed to consonantal voicing.

II. METHOD

The pharyngeal models considered in this study were constructed from MR images acquired while speakers of American English produced a range of fricative tokens in a scanner.

A. Subjects and corpora

Four monolingual native speakers of Standard American English, two women (W2 and W3) and two men (M1 and M2), were recruited as subjects. All were students of linguistics, aged between 21 and 26 years, who were paid for their participation. Non-naive subjects were deliberately chosen so that stimuli could be presented in the International Phonetic Alphabet (IPA), and the subjects could be instructed about the linguistic requirements of each task.

Each subject was asked to produce each of the English oral fricatives: eight consonants organized as voiced-voiceless pairs distributed over four places of articulation: labiodental [f]-[v], dental [θ]-[ð], alveolar [s]-[z], and post-alveolar [ʃ]-[ʒ]. Each fricative was elicited in three maximally distributed vocalic contexts [i_i], [a_a], and [u_u].

Stimuli were presented in the format [ifi if:i], allowing the subject to practice the token once, before taking a breath and sustaining the target fricative. The vowel immediately before the long fricative was sustained for approximately half a second before the scanner sequence was initiated at the beginning of frication. Subjects were instructed to sustain even frication and to concentrate on maintaining a consistent articulatory posture throughout the production. At the end of each sustained fricative, the final context vowel was repeated after the scanner had stopped to ensure that, as much as possible, the vocalic context had been maintained throughout the production.

Subjects practiced sustaining fricatives before each MRI session by producing the same corpora in an anechoic chamber while they were prompted using the same stimulus presentation software which was deployed in the scanner. High quality acoustic recordings were made during the anechoic sessions to allow for acoustic analysis of the speakers' fricatives.

A Siemens Sonata 1.5 T MRI scanner was used to image the subjects' vocal tracts while they produced all fricative tokens over the course of two 90-min sessions. Subjects lay supine in the scanner, sustaining each fricative in each vocalic context for 36 s. Prompts were presented in IPA, projected onto a screen which could be read by the subject from within the scanner bore. Subjects varied in the number of breaths they took during the sustained frication, from none (M1) to two (M2, depending on the trial); they were instructed to do so with a minimum of oral movement. In the few cases where the quality of an image sequence was compromised by motion blur associated with breathing or other activity, the token was reacquired.

All fricatives elicited during the scanning session were monitored using a FOMRI dual channel noise canceling optical microphone (Optoacoustics Ltd., 2007) integrated into the scanner. Scanner noise was attenuated by approximately 40 dB, allowing for real-time supervision of the subjects' speech. In this way the veracity of each fricative and vowel was monitored during the production of each token, as well as the maintenance of voicing in the long-hold fricatives. Scan sessions were organized such that, as much as possible, multiple tokens of the same fricatives and all tokens of voiced-voiceless fricative pairs were elicited in a single session to reduce alignment errors.

A two-dimensional True-FISP scan sequence (T_r = 200 ms, T_e = 3.3 ms, flip angle = 70°) was chosen as the best compromise between image resolution and scan time. Each token was repeated three times, so that it could be imaged in each of three orientations: sagittally (from ear to ear), axially (from upper trachea to nasal cavity), and obliquely, 45° to the axial planes, providing cross-sectional imaging of the tract in the alveolar region (from lips to velum) (see Proctor *et al.*, 2008 for details of imaging orientations). In each orientation, parallel slices of 4 mm thickness were acquired, spaced at 4.8 mm intervals. Although this interslice spacing was coarser than ideally desired, it was deemed more important to keep the scan time short to ensure that vowel context persisted through the production of each token. The number of slices acquired for each subject and each orientation was varied as needed in order to sample the tract over the region of interest.

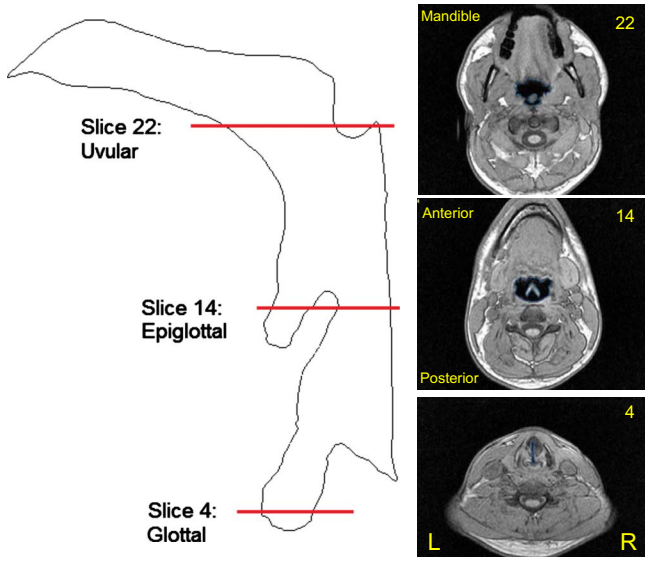


FIG. 1. (Color online) Left: Midsagittal outline of tract from glottis to alveolar constriction, showing location of three axial MRI slices for subject M1—[s] in token [as:a]. Right (bottom to top): Axial slice 4: through glottis; slice 14: midpharynx, showing epiglottis; slice 22: slice above top of oropharynx, showing tip of uvula.

B. Pharyngeal model construction

Initial processing of MR images was performed using 3D-DOCTOR, a vector-based three-dimensional image processing and modeling suite (Able Software Corp., 2007). All subsequent image processing and tract modeling was performed in MATLAB (MathWorks Inc., 2007). Image stacks were assembled from DICOM files in each orientation: sagittal, axial, and oblique. The MR images provided a resolution of 0.47×0.47 mm^2 in the plane of each image slice, with an interslice spacing of 4.8 mm.

For each fricative token, a subset of slices was selected from the axial stack to create a model of the subject's pharynx. A stack of 15–20 slices was required to cover the full extent of each oropharynx, depending on the size, gender, and anatomy of the subject. The bottom slice selected in each stack was the most constricted slice through the glottis above the elliptical sections defining the trachea (Fig. 1, slice 4). The top slice delineating the stack was chosen to be the last slice imaging the oropharynx before any evidence of the uvula was apparent (Fig. 1, slice 21). The sets of images chosen in this way corresponded to a 72–96 mm section of the pharynx.

Three-dimensional models were constructed from the subset of axial images defining the pharynx. Each image slice was segmented by identifying a set of points defining the air-tissue boundary, resulting in a set of curves lying in parallel planes representing the intersection of the tract with the center of each imaging plane. A model of the airway of each oropharynx was constructed from the axial stack of boundaries (Fig. 2). The tract model surface was created by fitting a triangulated mesh to the points defined on the con-

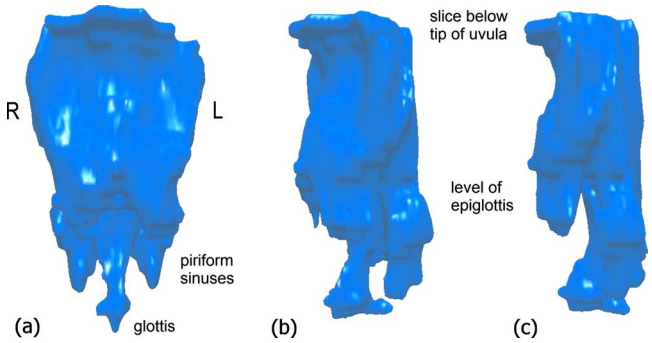


FIG. 2. (Color online) Pharyngeal model constructed from axial MRI slices—subject M1, fricative [z] in token [aza]. (a) Frontal view through mandible. (b) Front-left perspective. (c) Lateral view through left cheek.

stituent boundaries. The surface was triangulated using a DELAUNAY algorithm (de Berg, 1997).

C. Pharyngeal volume estimation

The volume of each pharynx was estimated from the triangulated surface defining the pharyngeal tract boundary. An arbitrary reference point on the tract surface was first chosen. For each triangle on the tract surface, a tetrahedron based on the triangle with an apex at the reference point was constructed, and the volume of the tetrahedron was calculated. All signed tetrahedral volumes defined over the surface with respect to the reference point were then summed to calculate the total volume enclosed by the surface.

III. RESULTS

A. Pharyngeal articulation and voicing

Comparison of the fricative tract models reveals that voiceless fricatives are generally produced with a different pharyngeal configuration to that employed in the production of voiced fricatives at the same place of articulation. Although the region of the tract around the fricative constriction is not shown in these models, no major differences were observed in the primary place of articulation between voiced and voiceless fricative pairs, nor in the size of the constriction.

Articulatory differences between voiced and voiceless fricatives are most apparent in the upper oropharynx, where voiceless tracts typically appear more constricted than their voiced equivalent tracts. Lateral views of the pharynx of subject W3 producing voiced and voiceless fricative pairs are shown in Fig. 3. Frontal views of the same pharyngeal models are shown in Fig. 4.

The vocal tract articulations employed in the production of the voiceless fricatives at all four places of articulation can be seen to be more constricted in the midpharyngeal region. The most constricted section of each of the voiceless pharynges is the region immediately above the epiglottis; in contrast, the voiced tracts maintain a more consistent volume throughout the upper pharynx.

The frontal views of the models (Fig. 4) reveal that the voiceless tracts are also more laterally constricted in the upper pharynx compared to their voiced equivalents, which are noticeably wider in the region below the uvula. For subject

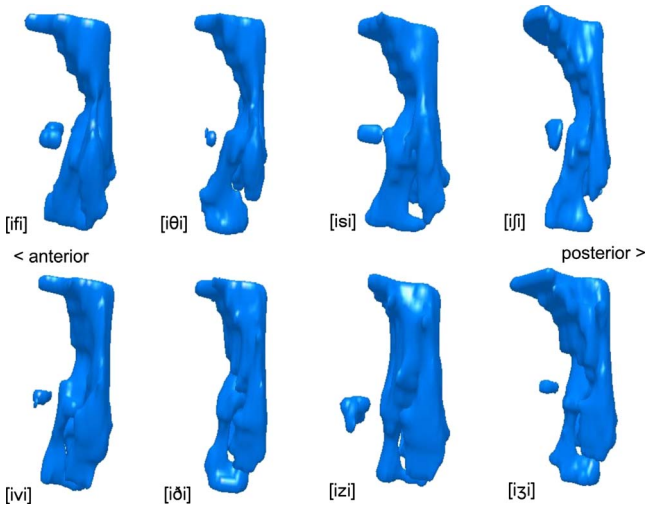


FIG. 3. (Color online) Pharyngeal models of voiceless (top row) and voiced (bottom row) fricatives. Left lateral view. Subject W3, all oral fricatives produced in [i_i] context.

W3, whose pharyngeal models are illustrated here, this effect is particularly noticeable in the sibilant tokens [izi] and [iʒi], which display considerable lateral expansion compared to their voiceless equivalents [isi] and [iʃi]. The issue of whether this expansion is passive or active is addressed in Sec. IV.

These same differences in pharyngeal articulation between voiced and voiceless fricatives are also evident in all tokens produced by subjects M2 and W2, across all four places of articulation. Subject M1 did not show the same consistent pharyngeal expansion for voiced tokens as the other subjects, and his pharyngeal articulation varied considerably with place of articulation. For some tokens, M1's pharynx was more expanded during the production of voiceless fricatives—an effect not observed in any other subject.

B. Pharyngeal volume and voicing

The gross differences in pharyngeal articulation observed in the tract surface models were quantified by calcu-

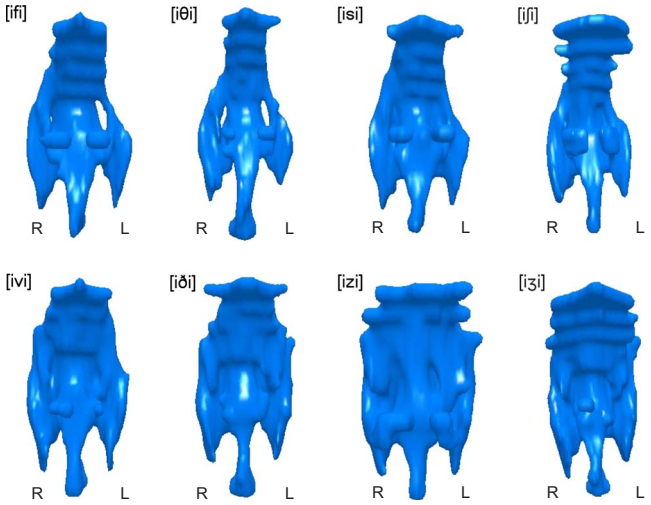


FIG. 4. (Color online) Pharyngeal models of voiceless (top row) and voiced (bottom row) fricatives. Frontal view. Subject W3, all oral fricatives produced in [i_i] context.

TABLE I. Mean pharyngeal volume (cm^3) by place of fricative articulation. All four subjects and all three vowel contexts. Percentage difference calculated as $(V_{voiced}-V_{unvoiced})/V_{unvoiced}\times 100$.

Place of articulation	Volume voiced	Volume unvoiced	Difference	
			cm^3	%
Labiodental	26.9	20.2	6.7	33
Dental	22.9	19.5	3.4	18
Alveolar	34.3	23.0	11.4	49
Postalveolar	33.1	23.7	9.3	39
All fricatives	29.3	21.6	7.7	36

TABLE II. Mean pharyngeal volume (cm^3) by subject. All fricatives and all vowel contexts. Percentage difference calculated as $(V_{voiced}-V_{unvoiced})/V_{unvoiced}\times 100$.

Subject	Volume voiced	Volume unvoiced	Difference	
			cm^3	%
M1	50.4	49.7	0.7	1
M2	37.9	18.9	19.0	100
W2	13.1	8.8	4.3	49
W3	15.8	9.0	6.8	76
All subjects	29.3	21.6	7.7	36

lating the volumes of the pharyngeal models for each token (Sec. II C). Differences in volumes of voiced and voiceless tracts were calculated for each pair of fricatives, and averaged over subjects and vowel contexts (Table I). Overall, for this group of speakers, voiced fricatives are produced with a 36% larger mean pharyngeal volume than their voiceless equivalents.

Mean pharyngeal volumes for each subject, averaged over fricative token and vocalic context, are given in Table II. The data show that three of these speakers produced voiced fricatives with a pharyngeal volume at least 49% larger than their voiceless equivalents. For one speaker (M2), the pharynx used to produce voiced fricatives is on average twice as large as the pharynx used to produce voiceless fricatives. Compared to the other subjects, the pharyngeal volume data of subject M1 were anomalous, showing no major difference in mean pharyngeal volume between voiced and voiceless fricatives.

Although mean voiced fricative pharyngeal volumes are consistently larger than voiceless volumes across all places of articulation, the magnitude and direction of these differences vary between subjects. Pharyngeal volumes of fricatives produced by individual subjects at each place of articulation are compared in Figs. 5–8.

The data show that three of the four subjects (M2, W2, and W3) consistently produce voiced fricatives with larger

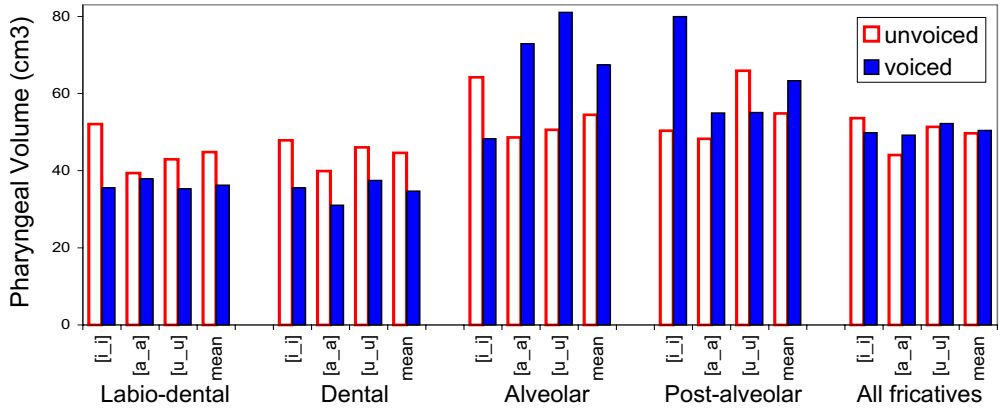


FIG. 5. (Color online) Fricative pharyngeal volumes by place of articulation and vocalic context—subject M1. Solid bars indicate voiced volumes; outlined bars indicate voiceless volumes. Contexts [i_i], [a_a], and [u_u] and mean volumes shown for each pair.

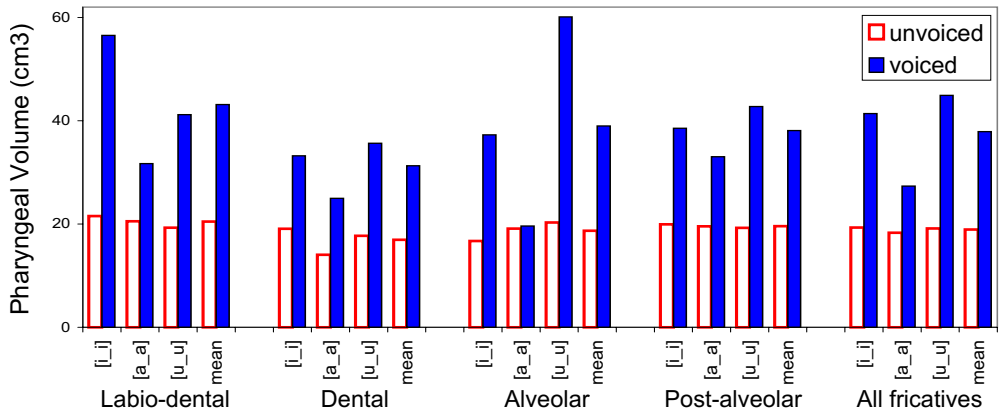


FIG. 6. (Color online) Fricative pharyngeal volumes by place of articulation and vocalic context—subject M2. See Fig. 5 for legend.

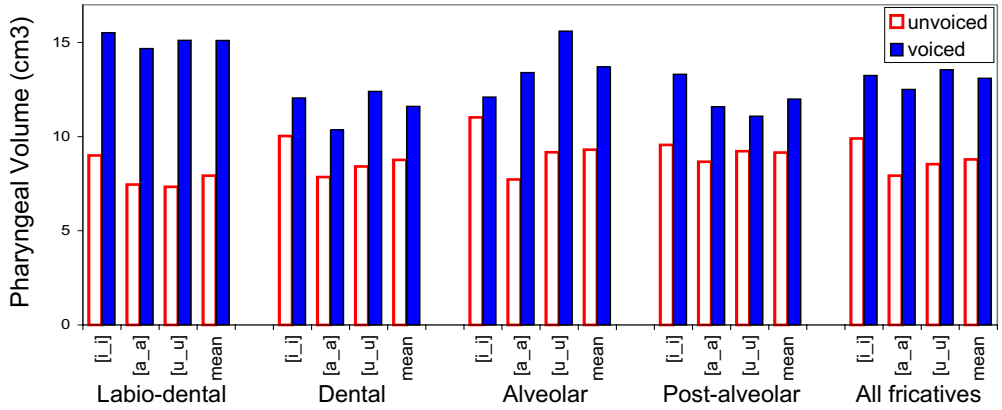


FIG. 7. (Color online) Fricative pharyngeal volumes by place of articulation and vocalic context—subject W2. See Fig. 5 for legend.

pharyngeal volumes than their voiceless equivalents, in all vocalic contexts at all places of articulation (Figs. 6–8). Subject M1 is anomalous in that his voiceless dental and labiodental fricatives, as well as the tokens [isi] and [uʃu], are all produced with a greater pharyngeal volume than that used in the production of their voiced fricative pairs (Fig. 5).

The data in Table II also reveal that the mean volume of the pharynx of the largest male subject (M1) was more than 3.5 times greater than the mean volume of the smallest female subject (W2). To provide a better means of comparison across the population of subjects, the data were normalized with respect to each subject.

Each pharyngeal volume was divided by the largest volume calculated for the same subject, resulting in four sets of relative volumes lying within the range [0–1]. The normalized pharyngeal volumes obtained in this manner were averaged over all subjects and grouped according to vowel context. The mean normalized volume data are illustrated in Fig. 9.

The data in Fig. 9 reveal that the fricative pair for which the voiced-voiceless pharyngeal volume differences are the most robust over all vowel contexts is [f-v]. For labiodental fricative tokens, the minimum differential pharyngeal volume is 49% (for vowel context [i_i]), the maximum differential volume 63% (vowel context [u_u]), and the mean difference across all contexts and speakers is 57%. The sibilants [s-z] also show a mean differential volume difference of 57%, but for this pair of fricatives the difference is less consistent, varying from 23% in the high front context [i_i] to 101% for context [u_u]. Although some of this variation is clearly attributable to the anomalous volume differences found in M1, subjects M2 and W2 also show large differences in the amount of pharyngeal volume change with sibilant voicing in different vowel contexts (compare M2 [isi-izi], [asa-aza], and [usu-uzu]).

C. Articulatory characterization of supraglottal changes during voicing

In order to examine the geometry of the pharynx during the production of voiced and voiceless fricatives in more detail, tissue outlines were extracted from selected MR imaging planes and superimposed to compare vocal tract configurations. Midsagittal images were superimposed to compare tract length, laryngeal position, and overall tract shape, and axial slices were used to examine cross-sectional differences in the epiglottal and pharyngeal regions.

1. Method

For each pair of voiced-voiceless fricative tokens, the two axial image stacks were aligned such that the configuration of the pharynx in each corresponding image slice could be compared across tracts. For each MR image stack, a Sobel edge-detection algorithm (Duda and Hart, 1973) was applied to automatically detect air-tissue boundaries. Contrast and edge-detection thresholds were selected to produce the best tissue outlines for the images of interest.

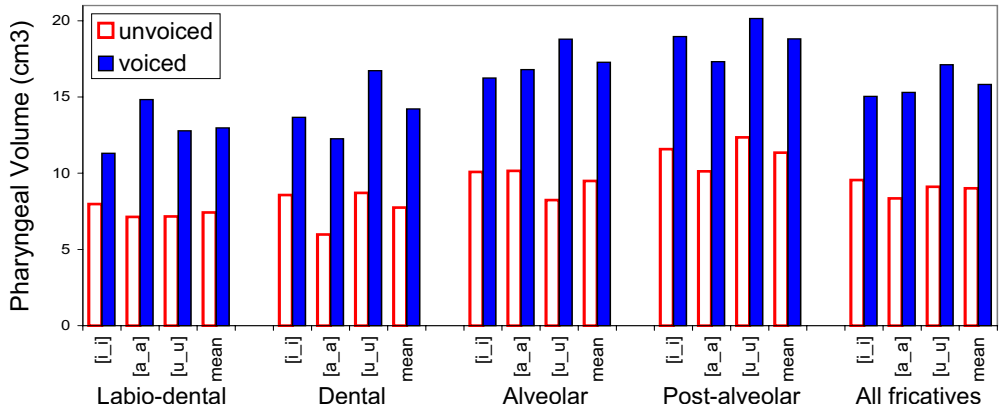


FIG. 8. (Color online) Fricative pharyngeal volumes by place of articulation and vocalic context—subject W3. See Fig. 5 for legend.

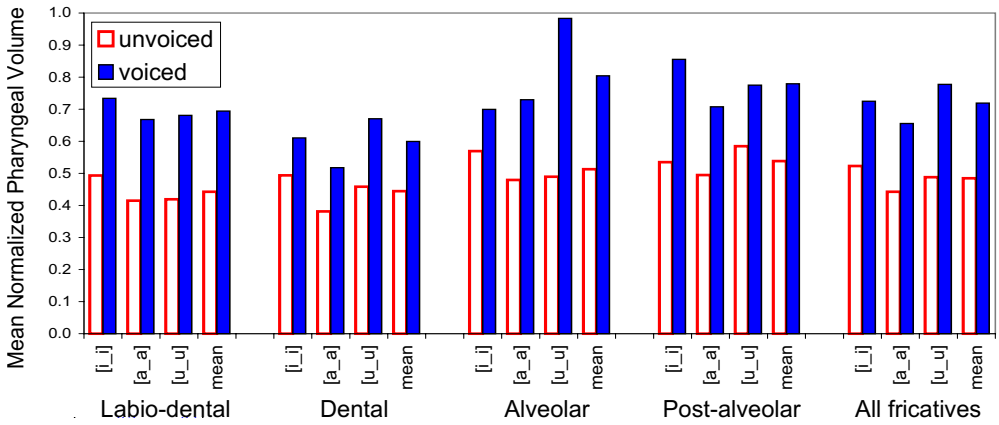


FIG. 9. (Color online) Mean normalized pharyngeal volumes by place of articulation and vowel context, averaged over all subjects. Solid bars indicate voiced volumes; outlined bars indicate voiceless volumes. Contexts [i_i], [a_a], and [u_u] and mean shown for each pair.

Tract boundaries corresponding to voiced and unvoiced tokens were superimposed. Alignment between image stacks was verified by comparing anatomical landmarks—primarily the outline of the subjects' head. In most cases, because the voiced-voiceless fricative pairs were acquired during the same imaging session, the subject's heads remained in perfect alignment, so that the superimposed tracts could be compared directly without any need for translation of the images with respect to each other. An example of a superimposed pair of image stacks generated in this way is shown in Fig. 10.

2. Characterizing pharyngeal differences

Area functions of fricative pairs were calculated from the axial image stacks in order to compare the contributions of different regions of the tract to the voiced-voiceless pharyngeal volume differences. For each slice in the image stack, the areas enclosed by the two tract boundaries were computed. A pair of area functions was constructed in this manner for each superimposed pharyngeal stack. Comparative area functions calculated from tract models of subject W3's productions of [uʃu] and [uʒu] are plotted in Fig. 11. Area functions comparing subject M2's tokens [usu] and [uzu] are shown in Fig. 12.

The area functions in Figs. 11 and 12 reveal that the bulk of the additional volume observed in the voiced pharynges is contributed by expansion of the upper pharynx. The same pattern is observed for all other tokens produced by subjects M2 and W3, and for all tokens produced by subject W2.

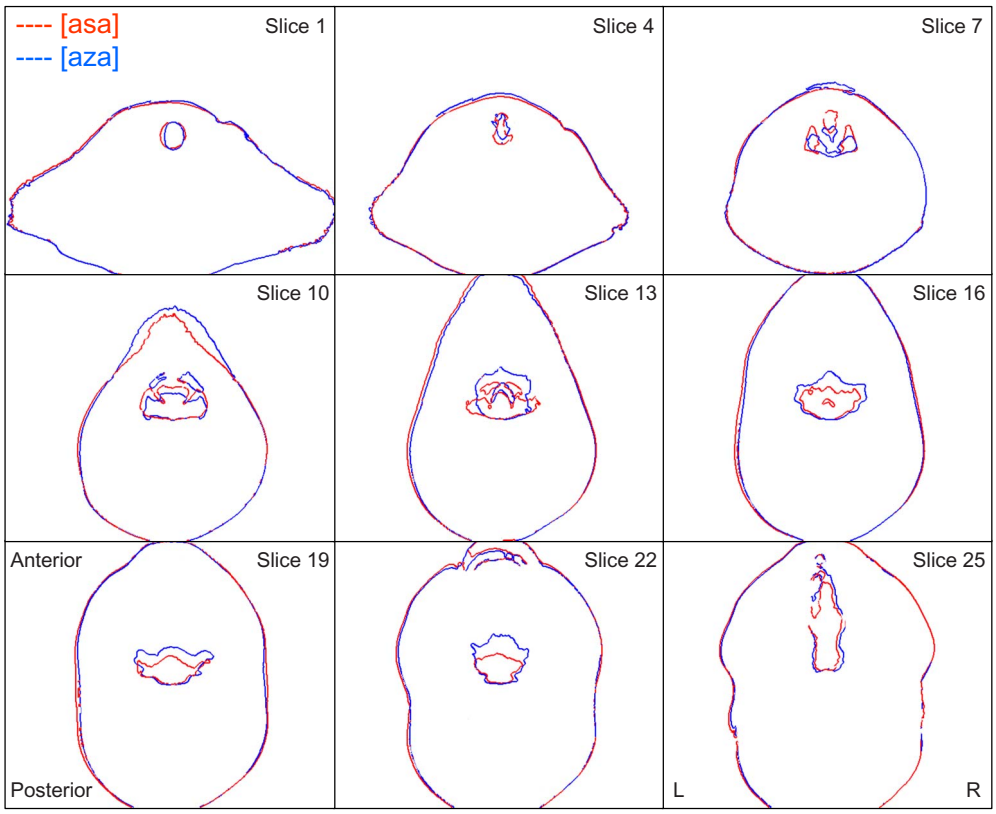


FIG. 10. (Color online) Superimposition of axial slices to compare pharyngeal articulation—subject M1, tokens [asa] vs [aza]. Higher slice numbers correspond to higher positions in tract. Slice 1 (top left): trachea; slice 7: supraglottal slice showing piriform sinuses; slice 13: midpharynx and epiglottis; slice 25 (bottom right): oral cavity beneath hard palate.

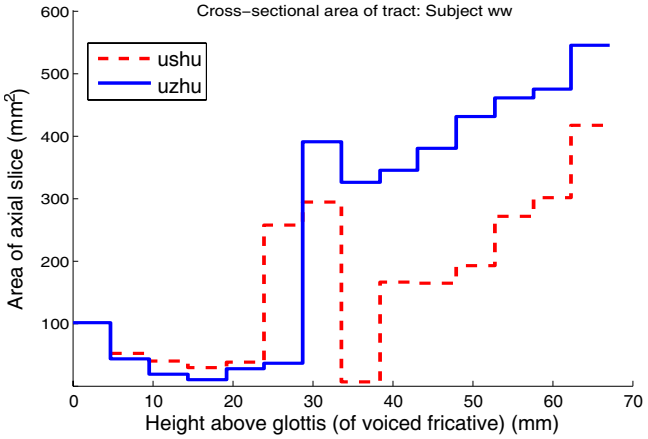


FIG. 11. (Color online) Comparative pharyngeal area function—subject W3, tokens [uʃu] (broken line) vs [uʒu] (solid line).

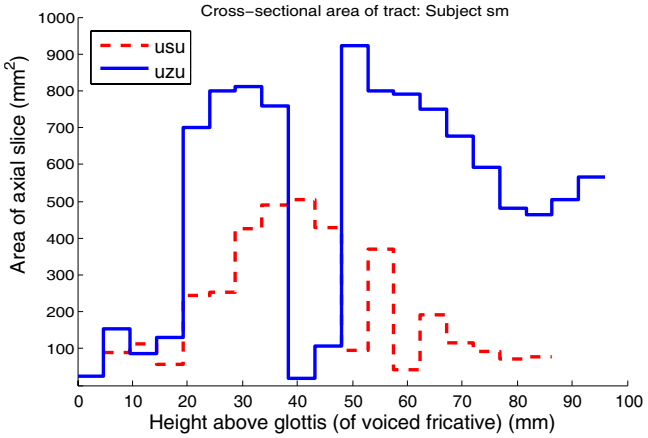


FIG. 12. (Color online) Comparative pharyngeal area function—subject M2, tokens [usu] (broken line) vs [uzu] (solid line).

The comparative area functions of fricatives produced by M1 also show the same pattern, except in those cases where the voiceless pharyngeal volumes are larger; in these cases—tokens [ifi], [ufu], [isi], and [uʃu]—the bulk of the additional volume in the voiceless pharynges is also contributed by expansion of the upper pharynx.

3. Pharyngeal variation throughout the pharynx

To gain further insights into the effect of vocalic context and place of articulation on fricative production, differential area functions were generated to observe the way in which the volume varies between voiced and voiceless productions in different parts of the tract. The data, organized by speaker and place of articulation, are illustrated in Fig. 13. Area functions are aligned by slice number within each subject such that slice 1 corresponds to the lowest position of the glottis for any token by that subject.

The differential area functions in Fig. 13 offer a number of insights into the difference between voiced and voiceless fricative pharyngeal articulation. Negative values indicate regions of the pharynx where the cross-sectional area of the voiced tract is greater than that used in the production of the corresponding voiceless fricative. The general pattern observed in all of these graphs is a function that starts at zero, becomes positive for 1–2 slices, and then trends negative for the remainder of the slices, corresponding to the upper half of the pharynx. Variations of this same general trend are observed for all subjects, including M1, whose upper laryngeal volume differences are anomalous compared to the other subjects. The fact that the differential area function alternates between positive and negative in this way indicates that the minimum pharyngeal areas occur at different heights for the same fricative place, vowel context, and subject.

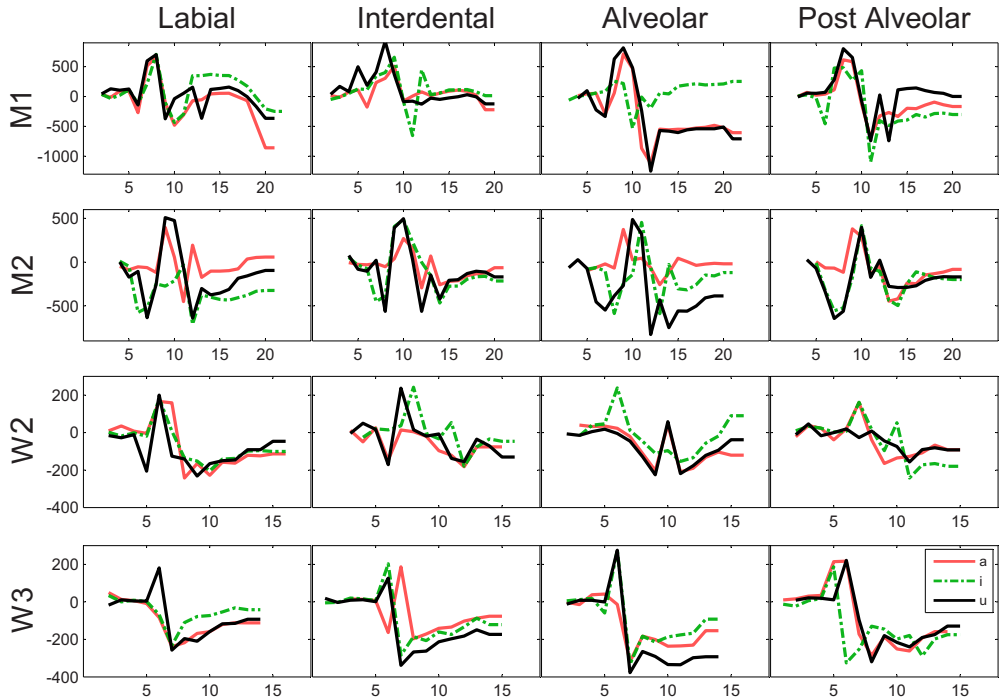


FIG. 13. (Color online) Differential area function ($A_{unvoiced}-A_{voiced}$, mm^2) by subject (row), place of articulation (column), and vocalic context: [i_i] (broken line), [a_a] (light solid line), and [u_u] (dark solid line). *x*-axis: image slice numbers. Higher slice numbers correspond to higher positions in pharynx.

The near-zero values at the beginning of the function correspond to the small volumes observed around the glottis in both voiced and voiceless fricatives. The functions trend rapidly positive and then negative because of differences in the vertical alignment of different parts of the larynx, piriform, and false vocal folds between the voiced and voiceless fricatives being compared, and because of differences in the volumes of these regions of the tract.

4. Laryngeal height

One of the important insights revealed by the area functions derived in Sec. III C 2 is that the glottal slices in the voiced fricative productions are often displaced with respect to the equivalent voiceless fricative productions. In both Figs. 11 and 12, for example, the area function of the voiced tract (solid line) begins one slice before the voiceless tract area function (broken line), indicating that the glottis is lower in the tract during the production of the voiced fricative.

In all 48 fricative pairs (4 subjects $\times$ 4 places of articulation $\times$ 3 vowel contexts), the height of the glottis during the production of the voiced fricative was either level with or up to three slices (14.4 ± 2.4 mm) lower in the tract than the voiceless fricative glottal height. Subject W3 showed the least glottal displacement with voicing, averaging 2.4 mm, and subject M2 the most (9.2 mm). The mean vertical glottal displacement across all 48 fricative pairs was 1.0625 slices or 5.1 ± 2.4 mm. That is, the glottis was on average half a centimeter lower in the voiced tracts than it was during the production of the voiceless fricative equivalents. This suggests that either the larynx is lowered during voicing or raised during the production of voiceless fricatives.

D. Mechanisms of expansion

Another question addressed in this study is whether the increased pharyngeal volume observed in the voiced fricative productions results from a uniform expansion of the upper pharynx or whether the additional volume is attributable to a particular active mechanism of articulation.

A comparison of midsagittal images of voiced-voiceless fricative pairs revealed no major differences in the height or configuration of the uvula, eliminating vertical displacement of the upper oropharynx as a mechanism of expansion for voiced production. For example, the tip of the uvula was located 59 mm below the top of the sphenoidal sinus during subject W3's production of [uvːu] and 58 mm below the sinus during her production of [ufːu]—a fricative pair which differs in volume by 43%, yet shows no evidence of vertical displacement of the uvula. Likewise, for subject M1, the uvular tip was measured to be 71 mm below the sphenoidal sinus during his production of [isːi] and 72 mm below the sinus during his production of [izːi].

In order to gain more insights into the nature of pharyngeal expansion, the voiced and voiceless tracts were compared by measuring the relative displacement of the front, rear, and side pharyngeal walls at different heights.

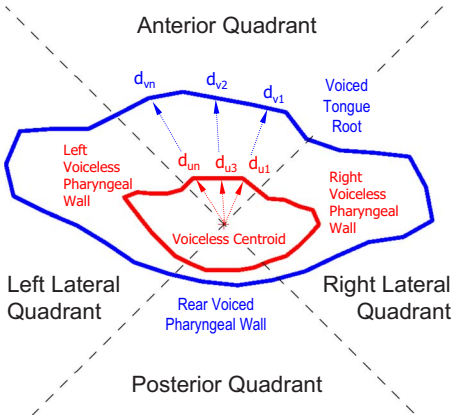


FIG. 14. (Color online) Superimposed voiced (outer) and voiceless (inner) pharyngeal boundaries, showing four quadrants and method of calculation of pharyngeal wall displacements in each quadrant—subject M2 [isi-izi] (slice 17).

In each slice of the pharyngeal stack, the centroid of the voiceless fricative tract was calculated. The image plane was divided into four quadrants by constructing two perpendicular axes intersecting at the centroid and oriented at 45° with respect to the anterior-posterior axis. The tract boundaries of the corresponding voiced fricative were superimposed on each slice, so that the relative articulations of the voiced-voiceless fricative pair could be compared. Because all fricative pairs were acquired during the same scanning session, no image registration was required to perform the superimposition, since each set of boundaries was represented within the same coordinate space.

1. Quantifying gross pharyngeal expansion

For both tract boundaries on each slice, the distance of each point to the voiceless centroid was calculated. By averaging both sets of distances in each quadrant, a mean displacement from the centroid was calculated for both voiced and voiceless tract boundaries. The relative displacement of the voiced tract with respect to the voiceless tract was then calculated as the difference between the mean boundary displacements (Fig. 14).

The method of calculation of the anterior pharyngeal differential distance A, for example, is given by

$$A = \frac{1}{n}\sum_{1}^{n} d_{Vi} - \frac{1}{m}\sum_{1}^{m} d_{Ui}, \tag{1}$$

where n, m = number of points on anterior quadrant of voiced and voiceless boundaries, respectively, and d is defined in Fig. 14. The posterior P, left lateral LL, and right lateral RL differential distances are calculated in the same way in their respective quadrants. The distance metrics calculated in this manner provide an indication of the amount of expansion of the voiced tract with respect to the voiceless tract at different heights in the pharynx.

In the lower pharynx, this pharyngeal expansion metric did not prove to be a reliable means of quantifying changes in articulation between voiced and voiceless fricatives because the morphology of this part of tract is not sufficiently simple. In the lower-midpharyngeal region, for example, the

TABLE III. Mean total differential pharyngeal wall displacement (dVoiced−dVoiceless) (mm), and ratios of anterior wall to total displacement and posterior wall to total displacement by subject—all fricatives and all vowel contexts.

Subject	Total V-U displacement (mm)	Anterior/total	Posterior/total
M1	3.19	0.15	0.15
M2	13.75	0.29	0.11
W2	8.47	0.51	−0.09
W3	11.65	0.29	0.08
Mean	9.27	0.31	0.06

epiglottis introduces a bifurcation, and in the laryngeal region, the tract trifurcates at the piriform sinuses. Another factor which prevents the use of the metric throughout the entire pharynx is the vertical displacement of the laryngeal region due to voicing (Sec. III C 4). This causes dissimilar parts of the pharynx to appear on the same image slice, which cannot be sensibly compared.

For these reasons, a subsection of each pharynx was chosen in which the voiced and voiceless tract morphologies were sufficiently alike to allow for comparison using this method—a section extending from the top slice in the pharynx down to the top of the epiglottis. The mean number of slices in the upper pharyngeal section was 4.85, representing a 23.3 mm section, covering approximately the top third of each pharynx.

In each slice of the upper pharynx, the total displacement of the voiced tract with respect to the voiceless tract was calculated as the sum of displacements in all four quadrants. The total pharyngeal displacement T was defined to be the sum of anterior, left lateral, right lateral, and posterior displacements [Eq. (2)].

$$T = A + LL + RL + P. \tag{2}$$

The gross differential pharyngeal wall displacement calculated using this metric varied considerably across fricative tokens, from a minimum of −9.30 mm (subject M1, tokens [ifi-ivi], reflecting the larger size of the voiceless tract) to a maximum of 31.05 (subject M2, tokens [usu-uzu]). The mean total differential upper pharyngeal wall displacement across all 48 fricative pairs was 9.27 mm (Table III).

2. Directional characterization of pharyngeal expansion

To better characterize the nature of pharyngeal expansion, the percentage contribution of pharyngeal wall displacement in each direction was estimated by calculating the ratio of directional displacement to total displacement T in each quadrant. The left (LL/T) and right (RL/T) displacement ratios were added to provide an indication of total lateral displacement (L/T), which could then be compared to the anterior (A/T) and posterior (P/T) displacement ratios. In each slice, by definition, the sum of anterior, posterior, and lateral displacement ratios is 1.

One pattern of upper pharyngeal expansion which was observed among these fricative pairs is shown in Fig. 15. The

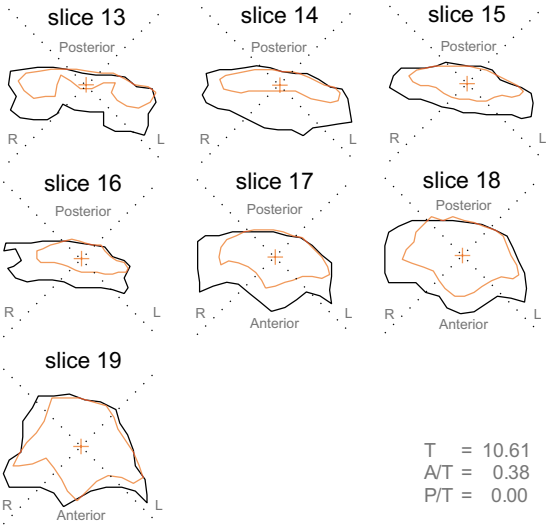


FIG. 15. (Color online) Upper pharyngeal expansion primarily through tongue root and lateral displacement—subject W2, upper eight slices of [ifi] (inner boundary) vs [ivi] (outer boundary). Slice 13: midpharynx, above epiglottis; slice 19: top of oropharynx, below uvula.

seven superimposed tract slices contrast subject W2's production of [ifi] (inner tract boundary) and [ivi] (outer boundary). The rear pharyngeal walls of both fricative productions are revealed to be in close alignment (P/T=0.00), and nearly all of the expansion of the voiced tract results from displacement in the lateral (L/T=0.59) and anterior (A/T=0.38) quadrants, suggesting that the sole mechanism of pharyngeal expansion employed in the production of the voiced fricative, in this case, involves forward displacement of the tongue root.

In a few cases, the expansion of the voiced pharynx was observed to occur more equally in all directions, as in the case of the fricative pair compared in Fig. 16. In this example (W2 [afa-ava]), 14% of displacement occurs in the posterior quadrant, 25% in the anterior quadrant, and the remainder is lateral (L/T=0.64). These data suggest that some volume differences might result from active constriction of the voiceless tract with a sphincterlike mechanism.

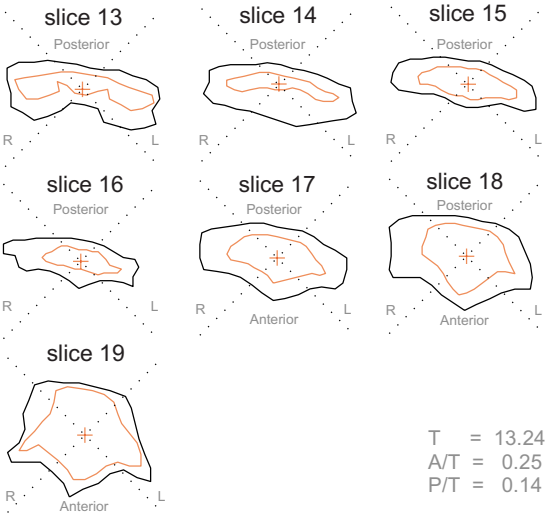


FIG. 16. (Color online) Concentric voiced/voiceless upper pharynges—subject W2, [afa] (inner boundary) vs [ava] (outer boundary). Slice 13: midpharynx, above epiglottis; slice 19: top of oropharynx, below uvula.

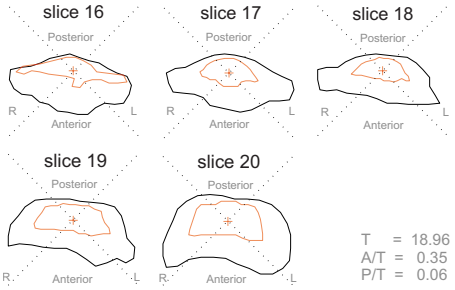


FIG. 17. (Color online) Typical displacement pattern observed in upper pharynx—subject M2, [isi] (inner boundary) vs [izi] (outer boundary). Slice 16: midpharynx, above epiglottis; slice 20: top of oropharynx, below uvula.

The alternative explanation—that the expansion of the voiced tract is passive—does not seem plausible because that would require increased intraoral air pressure in the voiced cases.

The prototypical pattern of pharyngeal expansion observed in this study involves neither a static rear pharyngeal wall (as in Fig. 15) nor a concentric arrangement of the voiced and voiceless tracts (as in Fig. 16), but rather the configuration illustrated in the fricative pair in Fig. 17. Some expansion of the voiced tract may be observed in all directions; however, the majority of the displacement is in the anterior and lateral quadrants, again suggesting that the primary mechanism of voiced pharyngeal expansion seems to involve forward displacement of the tongue root.

The extent to which forward movement of the tongue root dominates the expansion of the rear pharyngeal wall can be quantified by averaging the ratios of anterior and posterior pharyngeal wall to total displacement (Table III). For subject M1, no difference in anterior and posterior displacements was observed, a statistic which reflects the anomalous behavior of this subject in producing some voiceless fricatives with a larger pharynx than their voiced equivalents. However, for the remaining three subjects, the contribution of tongue root displacement to pharyngeal expansion (A/T) outweighs the effect of rear pharyngeal wall expansion (P/T) by a factor of at least 2. Overall, forward displacement accounts for 31% of total voiced expansion in the upper oropharynx.

IV. DISCUSSION

One issue to be considered when assessing the findings of this study is the artificiality of the task. MRI requires subjects to sustain fricatives for an unnaturally long time and to utter them while lying supine in a scanner.

Adopting a supine posture has been demonstrated to have a minor influence on articulation. Engwall (2003) observed slight pharyngeal narrowing in the supine production of Swedish vowels, compared to when prone, most noticeably for the front vowel [i]. In an X-ray microbeam study of two Japanese subjects, Tiede *et al.* (2000) concluded that the supine posture caused nonessential articulators to fall with gravity, while essential articulators are held in position even if against gravity.

Assuming these factors also applied to the subjects in this study, they should not affect the fundamental findings concerning difference between voiced and voiceless articulations, as all fricative tokens were produced under the same conditions. Furthermore Subtelny *et al.* (1972) observed the same direction of change between /s-z/ in a real-time study of subjects using upright posture, and Narayanan *et al.* (1995) observed a similar effect in the tongue root for subjects who were also in a supine position.

The sustained nature of the frication task asked of the subjects should be considered. In a study of Swedish vowels and fricatives, Engwall (2000, 2003) found the effect of sustaining long tokens to be equivalent in some respects to hyperarticulation. Tongue position was found to be more neutral for static holds (as opposed to dynamic), suggesting that the effects of coarticulation should be reduced in long holds.

The data presented in this study provide good evidence that coarticulatory effects persist in the subjects' productions of long-hold fricatives, since there is considerable variation in pharyngeal articulation due to vocalic context (as discussed in more detail in Shadle *et al.*, 2008). Although subjects may have used different strategies to manage airflow throughout the long holds, once more this should have affected both voiced and voiceless productions equally, so the insights into voicing differences remain valid. Factors associated with these different production strategies may explain some of the anomalies observed in the fricatives produced by subject M1, who was remarkable for having the largest lung capacity and perhaps needed to manage his airstream less than the other three subjects.

Given that the finding of a larger pharyngeal volume in voiced fricatives seems to be consistent with other studies, and is unlikely to be an artifact of the unnatural aspects of an MRI study, does Westbury's (1983) explanation for stops work for fricatives? First, Westbury (1983) showed that the vocal tract continued to expand during the closed phase of voiced stops. This does not hold for the voiced fricatives in the current study, as continual expansion would have resulted in blurred MR images. This does not in itself argue against Westbury's (1983) explanation, however; with the incomplete closure afforded by the fricative constriction, continued airflow through the glottis will not result in oral pressure increasing to the point where phonation would cease.

Is it possible, then, that the subjects find an equilibrium at the start of the long-sustained fricative where the air exiting through the constriction just balances the air entering through the phonating glottis, and the pharynx is expanded just enough to keep the pressure drop across the glottis in a range where phonation can continue? This explanation does not work either. The intraoral pressure must be less for voiced than for voiceless fricatives, unless subglottal pressure is sufficient to compensate for the pressure drop across the glottis in the voiced case. While this might be possible for sustained production in a MR scanner, that would be extremely unlikely to occur in normal speech, and thus would not explain Subtelny *et al.*'s (1972) results. With intraoral pressure lower in the voiced case, passive expansion due to air pressure would predict bigger pharyngeal volume in the voiceless case.

If not the result of passive expansion, then perhaps the expanded pharynx is the result of active expansion, held in one position throughout the long-sustained fricative. But it is

hard to come up with a convincing explanation for why this would make it easier, or more possible, to keep both voicing and frication sources going throughout. Instead, it seems possible that in the voiceless fricatives the tongue is actively pulled back in the upper pharynx to create a pressure-regulation mechanism. This could help reduce the airflow through the abducted glottis and would also help explain the anomalous results of subject M1: because of his greater lung capacity and large tract, subject M1 may not need to resort to this type of pressure regulation during long-hold frication as much as the other subjects.

V. CONCLUSIONS

In this study, the pharyngeal articulation of American English fricatives has been examined for four subjects using volumetric MRI data, and differences between voiced and voiceless fricatives have been characterized in three main respects.

The most important finding is that voiced fricatives are generally produced with a larger pharynx than that used during production of voiceless fricatives at the same place of articulation. These volume differences are consistent across all places of articulation and vocalic contexts for three of the four subjects studied. The subject with the largest pharynx differed from the other four by producing labiodental and interdental voiceless fricatives with a larger pharyngeal volume.

The bulk of the additional volume observed in the voiced fricatives was found to result from the expansion of the upper pharyngeal region above the epiglottis, below the uvula. The chief mechanism of expansion of the upper oropharynx was found to be displacement of the tongue dorsum—which was estimated to contribute approximately 31% of the additional volume—and the lateral pharyngeal walls, rather than through displacement of rear pharyngeal walls (6% of additional volume). Although the displacement of the rear wall was found to be smaller than that of the anterior portion of the pharynx, these findings also suggest that the posterior pharyngeal wall cannot simply be regarded as an immovable part of the vocal tract, as has been suggested.

Finally, the larynx was found to be consistently lower during production of voiced fricatives than during production of voiceless fricatives.

Many of the differences between voiced and voiceless fricatives observed in this study were only discovered by examining the geometry of the pharynx in all three dimensions. This demonstrates that midsagittal analysis of the vocal tract alone is insufficient to properly characterize fricative production and voicing.

This study has addressed the broad pharyngeal differences which can be observed in the production of voiced and voiceless fricatives; however, there are many more aspects of fricative production and voicing which remain to be investigated. More phonetic data are required to reconcile the differences in volume observed here with more detailed aspects of mechanisms of voicing and frication, and to account for some of the inconsistencies observed between subjects, and at different places of articulation.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, Grant No. NIH-NIDCD-RO1-DC006705. The authors are grateful to Carol Fowler, Christine Mooshammer, and two anonymous reviewers for their comments on an earlier version of this article.

Able Software Corp. (**2007**). 3D-DOCTOR version 3.5, Able Software Corp., Lexington, MA.

Bell-Berti, F. (**1975**). "Control of pharyngeal cavity size for English voiced and voiceless stops," J. Acoust. Soc. Am. **57**, 456–461.

de Berg, M. (**1997**). *Computational Geometry: Algorithms and Applications* (Springer, Berlin).

Duda, R. O., and Hart, P. E. (**1973**). *Pattern Classification and Scene Analysis* (Wiley, New York).

Engwall, O. (**2000**). "Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG, and EMA," in Proceedings of the International Conference on Spoken Language Processing, Vol. **I**, pp. 17–20.

Engwall, O. (**2003**). "A revisit to the application of MRI to the analysis of speech production—Testing our assumptions," in Proceedings of the 6th Seminar on Speech Production, Sydney, Australia, pp. 43–48.

Hardcastle, W. J. (**1976**). *Physiology of Speech Production: An Introduction for Speech Scientists* (Academic, London).

Kent, R. D., and Moll, K. L. (**1969**). "Vocal-tract characteristics of the stop cognates," J. Acoust. Soc. Am. **46**, 1549–1555.

Ladefoged, P. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech," J. Acoust. Soc. Am. **35**, 454–460.

Magen, H. S., Kang, A. M., Tiede, M. K., and Whalen, D. H. (**2003**). "Posterior pharyngeal wall position in the production of speech," J. Speech Lang. Hear. Res. **46**, 241–251.

MathWorks Inc. (**2007**). MATLAB version R2007b, MathWorks Inc., Natick, MA.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325–1347.

Ohala, J. J. (**1983**). "The origin of sound patterns in vocal tract constraints, *The Production of Speech* (Springer-Verlag, New York), Chap. 9, pp. 189–216.

Optoacoustics Ltd. (**2007**). FOMRI-II version 2.2, Optoacoustics Ltd., Or-Yehuda Israel.

Perkell, J. S. (**1965**). "Studies of the dynamics of speech production," Quarterly Progress Report, Mass. Inst. Tech. Res. Lab. Elect. **76**, 253–257.

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT, Cambridge, MA).

Proctor, M. I., Shadle, C. H., and Iskarous, K. (**2008**). "A method of co-registering multiple magnetic resonance imaged vocal tract volumes for fricatives," in Proceedings of the Joint Meeting of the Acoustical Society of America and European Acoustics Association, Paris, France, pp. 5093–5098.

Rothenberg, M. (**1968**). *The Breath-Stream Dynamics of Simple-Released-Plosive Production*, Biblioteca Phonetica Vol. **6** (Karger, Basel).

Shadle, C. H., Proctor, M. I., and Iskarous, K. (**2008**). "An MRI study of the effect of vowel context on English fricatives," in Proceedings of the Joint Meeting of the Acoustical Society of America and European Acoustics Association, Paris, France, pp. 5099–5104.

Stetson, R. H. (**1951**). *Motor Phonetics: A Study of Speech Movements in Action* (North-Holland, Amsterdam).

Subtelny, J. D., Oya, N., and Subtelny, J. D. (**1972**). "Cineradiographic study of sibilants," Folia Phoniatrica (Basel) **24**, 30–50.

Tiede, M. K., Masaki, S., and Vatikiotis-Bateson, E. (**2000**). "Contrasts in speech articulation observed in sitting and supine conditions," in Proceedings of the 5th Seminar on Speech Production, Kloster Seeon, pp. 25–28.

Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**, 1322–1336.

Influence of the ventricular folds on a voice source with specified vocal fold motion[a)]

Richard S. McGowan[b)]
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420

Michael S. Howe
College of Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

(Received 11 December 2008; revised 30 June 2009; accepted 2 July 2009)

The unsteady drag on the vocal folds is the major source of sound during voiced speech. The drag force is caused by vortex shedding from the vocal folds. The influence of the ventricular folds (i.e., the "false" vocal folds that protrude into the vocal tract a short distance downstream of the glottis) on the drag and the voice source are examined in this paper by means of a theoretical model involving vortex sheets in a two-dimensional geometry. The effect of the ventricular folds on the output acoustic pressure is found to be small when the movement of the vocal folds is prescribed. It is argued that the effect remains small when fluid-structure interactions account for vocal fold movement. These conclusions can be justified mathematically when the characteristic time scale for change in the velocity of the glottal jet is large compared to the time it takes for a vortex disturbance to be convected through the vocal fold and ventricular fold region.
 [DOI: 10.1121/1.3299200]

PACS number(s): 43.70.Bk, 43.70.Gr [DOS] Pages: 1519–1527

I. INTRODUCTION

The larynx is a complex structure of muscle, cartilage, and epithelium where voiced sounds can be created. The vocal folds consist of a pair of layered structures that partially obstruct the flow produced when the lung pressure is somewhat greater than the pharyngeal pressure. Muscles adjust the positions of the folds to permit flow-structure interaction that drives the vocal folds into oscillation. These oscillations periodically change both the size of the region of air between the folds, known as the glottis, and the resistance to flow through the glottis (van den Berg *et al.*, 1957; Fant, 1960). The resistance to flow is equal and opposite to the drag experienced by the vocal folds, and the resulting fluctuations in the volume velocity of air into the vocal tract generate sound that is heard to emanate from the mouth (Titze, 1994; Stevens, 1998). The sound coming from the laryngeal region can be viewed as having been generated from a fluctuating drag, or pressure expansion waves, superposed on an otherwise uniform pressure wave traveling from the lungs (Zhao *et al.*, 2002; Howe and McGowan, 2007).

Howe and McGowan (2007) presented a theory of the voice source derived using an aeroacoustic analogy and the method of compact Green's functions (Howe, 1975, 1998, 2003). They showed that the major source of sound, the fluctuating drag on the vocal folds, can be calculated with knowledge of the distribution of fluid velocity in regions where vorticity is nonzero, which can be idealized to be a very thin shear layer at the boundary between the glottal jet and the surrounding air in the vocal tract. The calculation of the magnitude of the drag or, equivalently, the source's effect on the output acoustics is localized to be within the shear layer just downstream of the glottal exit because of the shape of the vocal folds. The convolutory solution with a compact Green's function makes explicit the required alignment between the fluid velocity field near the vocal folds and the shape of the vocal folds in order that there be nonzero drag. The effect of the shape of the vocal folds on tissue-air forces is expressed in a factor of the Green's function known as the *Kirchhoff vector*. Because drag is the force of interest, it is the axial component of the Kirchhoff vector that is relevant. Roughly, the axial component of the Kirchhoff vector for the vocal folds is the velocity potential field with unit inflow and outflow in the direction of the vocal tract axis—it depends only on the shape of the tissue boundaries. This potential flow has a stream function, which is here termed the *axial Kirchhoff stream function*. Figure 1(a) shows model vocal folds and Fig. 1(b) shows the axial Kirchhoff streamlines, on which the axial Kirchhoff stream function is constant, below the vocal tract axis of symmetry. At the glottal exit region the gradient of axial Kirchhoff stream function has a large component parallel to the glottal jet velocity; this is a necessary condition for the generation of drag between the tissue and air (e.g., Howe, 1998, 2003).

Howe and McGowan (2007) based their analysis on a relatively simple geometry in order to derive drag on the vocal folds: vocal folds in a uniform duct. The present paper extends that account to tissue structure just downstream of the vocal folds, known as the ventricular folds, or false vocal folds. Like the vocal folds, they protrude into the vocal tract at a distance generally less than 1 cm from the downstream end of the vocal folds (Agarwal *et al.*, 2003). The streamlines of the axial component Kirchhoff vector will certainly

[a)]This paper is based on work presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: rsmcgowan@cressllc.net

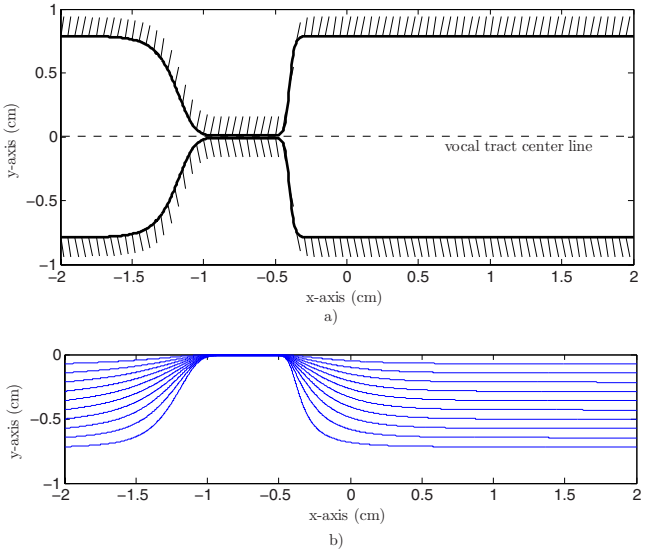


FIG. 1. (Color online) (a) Model vocal folds in the midsagittal plane and (b) axial Kirchhoff stream function streamlines below the axis of symmetry.

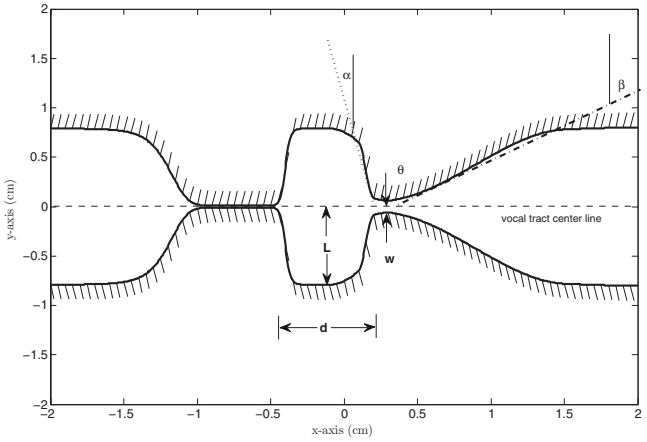


FIG. 2. Model vocal folds and ventricular folds in the midsagittal plane.

be altered near the ventricular folds compared to when they are assumed to be absent. Further, the ventricular folds are proximate enough to the vocal folds that the streamline patterns near the glottal exit can be expected to be altered as well. Either can produce change in the overall force between the tissues and the air near the vocal folds and, thus, affect the radiated acoustic field.

Zhang *et al.* (2002) provided a calculation of the effect of the ventricular folds on the voice source based on computational fluid dynamics (CFD). They found little effect of the ventricular folds on the voice source below 2000 Hz. Routine use of CFD can be very expensive and time-consuming, and, thus, it is difficult to confirm predictions for many possible fundamental frequencies and ventricular fold shapes. The approach of this paper is to generalize the analytic results of Howe and McGowan (2007) to include obstructions close to the vocal folds, which also requires some computational work, but at a much lower effort than full scale CFD.

The movement of the vocal folds will be specified by a simple mathematical model, and the ventricular folds are assumed to be rigid. This kind of analysis is employed when the effects of changes in fluid mechanics are considered without the complications of changes in vibrational patterns caused by fluid-structure interaction (e.g., Zhang *et al.*, 2002). These interactions are a part of the aerodynamic-myoelastic theory of vocal fold vibration. Therefore the results of our study are limited to the determination of the effect of the ventricular folds on the air flow itself, without quantitative consideration on how the ventricular folds influence the vibration of the vocal folds. Section II reviews some of the basic theory from Howe and McGowan (2007) and concludes with an equation for calculation of the drag if structures close to the vocal folds are included. Section III introduces the approximations used in the numerical calculation of the output pressure variation with and without ventricular folds. Section IV presents the results of these calculations, showing that the ventricular folds are expected to have a small effect on the propagated pressure wave from the glottis at frequencies below 2 kHz. In Sec. V it is shown that this is the expected result given the behavior of the axial component of the Kirchhoff vector and the quasi-steady nature of the low-frequency portions of the glottal velocity pulse.

II. THEORETICAL BACKGROUND

The geometries of concern are illustrated in Figs. 1(a) and 2. In Fig. 1(a) the vocal folds form a constriction, the glottis, and in Fig. 2 vocal folds are shown forming a constriction to the left (upstream), and the ventricular folds form a secondary constriction to the right (downstream) in a rigid duct of uniform width $2h$. The figures show a two-dimensional geometry in the x-y plane, with the z direction orthogonal to this plane. The x-axis is the duct centerline and it denotes the axial direction. The coordinate system origin is taken to be about 0.4 cm downstream from the glottis in its rest state. The distance between the downstream end of the vocal folds and ventricular fold constriction is of the order of 1 cm, and, together, these structures occupy less than 4 cm of the vocal tract axis. The source region including both structures can therefore be considered to be acoustically compact for frequencies below about 5 kHz. Only plane wave propagation is assumed to occur in the duct.

Peak air speeds in the glottis during phonation at a conversational level should be less than 4000 cm/s. Assuming that the amount of heat created by viscous action in the vocal tract is small and body temperature does not vary substantially, the flow can be considered to be homentropic. Howe's (1998) formulation of Lighthill's (1952) equation for the production and propagation of sound at low Mach number flow that may be considered homentropic is

$$\left(\frac{1}{c_o^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)B = \mathrm{div}(\boldsymbol{\omega} \wedge \mathbf{v}), \tag{1}$$

where c_o is the mean speed of sound, which is approximately uniform and constant in low speed flows, and B is the *total enthalpy* defined in homentropic flow by

$$B = \int \frac{dp}{\rho} + \frac{1}{2}v^2, \tag{2}$$

where $v=|\mathbf{v}|$, $\mathbf{v}$ is the fluid velocity, $\boldsymbol{\omega}=\mathrm{curl}\mathbf{v}$, the vorticity, and $\rho\equiv\rho(p)$ is the density.

Let the incident over-pressure from the lungs consist of a step rise in pressure of amplitude p_I impinging on the glottis at $t=0$. Then the corresponding incident total enthalpy wave B_I can be written as

$$B_I = B_I\left(t-\frac{x}{c_o}\right)=\frac{p_I}{\rho_o}H\left(t-\frac{x}{c_o}\right), \tag{3}$$

where H is the Heaviside step function. Put $B=B_I+B_s$, where B_s is the field scattered at the vocal folds, and the ventricular folds if they are present, which has outgoing wave behavior at large distances from the glottis.

Howe and McGowan (2007) derived the following integral representation of the solution to Eqs. (1) and (3):

$$\begin{aligned} B(\mathbf{x},t) = \frac{p_I}{\rho_o}\Bigg\{ & H\left(t-\frac{x}{c_o}\right) \\ & -\frac{\mathrm{sgn}(x)}{\ell([t])}\int_{-\infty}^{[t]}\frac{\partial}{\partial\tau}(H(\tau)\ell(\tau))e^{-\int_\tau^{[t]}2c_o d\xi/\ell(\xi)}d\tau\Bigg\} \\ & -\int[(\boldsymbol{\omega}\wedge\mathbf{v})(\mathbf{y},\tau)]\cdot[\nabla_{\mathbf{y}}G(\mathbf{x},\mathbf{y},t,\tau)]d^3\mathbf{y}d\tau \\ & +\oint_{S(\mathbf{y},\tau)}G(\mathbf{x},\mathbf{y},t,\tau)\left[\frac{\partial\mathbf{v}}{\partial\tau}(\mathbf{y},\tau)\right]\cdot d\mathbf{S}(\mathbf{y})d\tau, \end{aligned} \tag{4}$$

where G is the acoustic Green's function associated with the geometry of either Fig. 1(a) or Fig. 2, $[t]=t-|x|/c_o$ is the acoustic delay time, and viscous terms have been neglected, which is valid for high Reynolds number flows, and thus, for all times except possibly when the vocal folds are nearly closed. $\ell(t)$ is the time-dependent Rayleigh "end correction" of the glottal constriction and ventricular fold constriction, if it exists (Rayleigh, 1945). [Note that unprimed coordinates will be associated with the field coordinate vector $\mathbf{x}$ and primed coordinates with the source coordinate vector $\mathbf{y}$. Thus, $\mathbf{x}=(x,y,z)$ and $\mathbf{y}=(x',y',z')$.]

Let Y be the axial Kirchhoff vector at time t for either the geometry without ventricular folds [Fig. 1(a)] or with ventricular folds (Fig. 2). It is a solution of Laplace's equation satisfying $\partial Y/\partial x_n'=0$ on the surfaces $S(t)$ of the folds and the duct walls, and can be normalized such that

$$Y\sim x'\pm\frac{\ell(t)}{2},\quad x'\to\pm\infty. \tag{5}$$

The end correction is given by

$$\ell(t)=\int_{-\infty}^{\infty}\left(\frac{\partial Y}{\partial x'}(\mathbf{y},t)-1\right)dx', \tag{6}$$

provided the duct cross-sections are the same on both sides of the region containing the vocal folds and ventricular folds, if they exist. The integration is along any path parallel to the duct axis passing through the constriction(s).

The method of compact Green's function was used to derive an approximation to the Green's function appearing in Eq. (4) (Howe and McGowan, 2007). The compact approximation to the Green's function for source positions in the vicinity of the constrictions, $\mathbf{y}\sim h$, is

$$\begin{aligned} G(\mathbf{x},\mathbf{y},t,\tau)\approx & \frac{c_o}{2\mathcal{A}}H([t]-\tau)+\frac{c_o\,\mathrm{sgn}(x)Y(\mathbf{y},\tau)}{\ell([t])\mathcal{A}} \\ & \times H([t]-\tau)e^{-\int_\tau^{[t]}2c_o d\xi/\ell(\xi)},\quad \mathbf{y}\sim O(h), \end{aligned} \tag{7}$$

where $\mathcal{A}=2h\ell_3$ is the cross-sectional area of the duct and ℓ_3 is the length in the z-dimension. The differences between Green's function in a geometry without ventricular folds [Fig. 1(a)] and a geometry with ventricular folds (Fig. 2) are governed by the corresponding differences in the axial component of the Kirchhoff vector Y.

The first term of Eq. (4) is the incident pressure wave, the second term is the initial transient response, the third term is the vortex sound source or drag term, and the fourth term represents the sound created by changes in the volume of the vocal folds. In this paper we are concerned with the evaluation of the vortex sound when ventricular folds are placed just downstream of the vocal folds. The fourth term, which is the monopole strength caused by unsteady volume change, will not be considered further because the absence or presence of the ventricular folds has no effect on this term.

After the step wave of pressure given in Eq. (3) reaches the vocal folds, vorticity is shed from the vocal folds and it is convected downstream in jet shear layers. The strength of the vorticity shed from the vocal folds depends on the relative velocity between the jet and the air outside the jet and, to a first approximation, is proportional to the velocity of the jet. Further, neglecting the effects of viscous dissipation, the circulation of each vortex element of the shed vorticity remains constant as it is convected downstream. The typical approximation is that the shear layer vorticity convects at one-half the local jet velocity that is associated with its circulation. Vortex sound is directly related to the fluctuating drag on the structures in the duct (i.e., forces on structures directed along the axis of the duct).

If the far-field (i.e., typically at a distance from the source region exceeding a duct diameter) is considered and the contribution from the fluctuating volume term (the fourth term on the right) in Eq. (4) is excluded, $B(\mathbf{x},t)\approx p'(x,t)/\rho_0$, where the prime denotes the pressure field without the monopole term. For $x>0$ the acoustic particle velocity of this outgoing wave is equal to $p'(x,t)/\rho_o c_o$. Let

$$U(t)=\lim_{x\to+0}\frac{p'(x,t)}{\rho_o c_o} \tag{8}$$

be the limiting value of this velocity just to the right of the vocal fold region in Fig. 1(a) or the vocal fold and ventricular fold region in Fig. 2. Also, $x\sim+0$ so that $[t]\to t$. In this limit, an equation for $U(t)$ can be derived in the form

$$\frac{d(\ell U)}{dt}+2c_o U+B_\omega=\frac{2p_I}{\rho_0}H(t), \tag{9}$$

with

$$B_\omega = \int [(\boldsymbol{\omega} \wedge \mathbf{v})(\mathbf{y},\tau)] \cdot [\nabla_{\mathbf{y}} G(\mathbf{x},\mathbf{y},t,\tau)] d^3\mathbf{y} d\tau, \qquad (10)$$

where B_ω is the vortex sound term from Eq. (4). Equation (9) is a modification of one found in Howe and McGowan, 2007 [Eq. 6.3] with B_ω left unevaluated so that the effect of geometric changes in the laryngeal region, especially the inclusion of ventricular folds, can be studied. In Howe and McGowan, 2007 B_ω was evaluated in terms of jet velocity using the fact that the jet was narrow compared to the width of the duct and without further obstructions downstream. With the addition of ventricular folds, a numerical procedure is necessary to evaluate this integral.

III. METHOD

A. Vocal folds and ventricular folds

The effect of the ventricular folds on the acoustic output was determined by solving Eqs. (9) and (10) numerically. The geometry of the vocal folds was based on a model for the vocal folds given by Zhao *et al.* (2002). The model for the vocal fold movement had two-dimensional symmetry and was adapted from the original cylindrical geometry (Howe and McGowan, 2007). A further modification was made here so that the folds were narrower in the axial direction and the faces of the vocal folds were steeper, to make them accord with known dimensions. Let $\Delta(x,t)$ be the distance between the vocal folds, so that

$$\Delta(x,t) = \tfrac{1}{2}(D_0 + D_{\min} + (D_0 - D_{\min})\tanh(\hat{s})) + D_{\max}[1 - \tanh(\hat{s})]((\hat{x}+\hat{c})\beta_1(T) - (\hat{x}-\hat{c})\beta_2(T)), \qquad (11)$$

where $T = f_o t - [f_o t]$ is the fractional part of $f_o t$, and where

$$\hat{s} = \hat{b}\left(|\hat{x}| - \frac{1}{|\hat{x}|}\right),$$

$$\hat{x} = \frac{x + 0.8}{D_{\max}},$$

$$\hat{b} = \begin{cases} 1.4 & \text{if } \hat{x} < 0 \\ 6.0 & \text{otherwise,} \end{cases}$$

$$\hat{c} = 0.42,$$

$$\beta_1(T) = \begin{cases} 0 & \text{if } T \le \frac{1}{9} \\ 0.244\left\{1 - \cos\left[\frac{9\pi}{4}\left(T - \frac{1}{9}\right)\right]\right\} & \text{if } \frac{1}{9} < T \le \frac{5}{9} \\ 0.488 & \text{if } \frac{5}{9} < T \le \frac{6}{9} \\ 0.244\left\{1 + \cos\left[3\pi\left(T - \frac{6}{9}\right)\right]\right\} & \text{if } \frac{6}{9} < T \le 1, \end{cases}$$

$$\beta_2(T) = \beta_1\left(T + \tfrac{1}{9}\right),$$

$\Delta_0 = h$, $D_{\min} = 0.02$ cm, and $D_{\max} = 0.2$ cm. With these parameters the minimum glottal width $\Delta_m(t)$ varied from 0.015 to 0.1 cm.

Data provided in Agarwal *et al.*, 2003 were used to guide the choice of the dimensions of the ventricular folds and their position relative to the vocal folds for the numerical experiments. These measures were taken from laminagraphic images while subjects phonated, and those used in this study are illustrated in Fig. 2. The only modification to the dimensions of Agarwal *et al.* (2003) was to make the distance between the ventricular folds smaller than measured. It is possible for human speakers to reduce this distance, and we wanted to maximize the possibility that the ventricular folds could have any effect on the output acoustics. All of the measured values refer to Agarwal *et al.* (2003). The gap between the ventricular folds in measurements was between 0.2 and 0.8 cm, so the distance from a ventricular fold to the vocal tract axis w was measured to be between 0.1 and 0.4 cm. We have taken the distance w to be 0.06 cm in this work, which meant that the height of the lower vocal fold at minimum fold separation was only 0.01 cm higher than the lower ventricular fold when the glottis was maximally open. The measured distance d from the downstream edge of the vocal folds to the false folds ranged from 0.2 to 0.75 cm, and the distance used here was 0.7 cm. The measured maximum width of the channel between the ventricular sinuses ranged from 0.9 to 2.3 cm, and, thus, the maximum distance L from the bottom of one ventricular sinus to the vocal tract axis was between 0.45 and 1.15 cm, while $L = 0.8$ cm here. (The ventricular sinus is the gap downstream of the vocal folds and upstream of the ventricular folds.) The measured interior angle of the lines tangent to the downstream face and upstream face of the ventricular folds θ ranged between 55° and 107°. In this paper we let $\theta = 80°$. The angle α between the tangent of the upstream ventricular fold surface and the perpendicular to the vocal tract axis was measured to be between −18° and 51°, and $\alpha = 15°$ in this paper. Finally the measured angle between the tangent of the downstream ventricular fold surface and the perpendicular to the vocal tract axis β was between 22° and 76°, and $\beta = 65°$ in this work.

B. Numerical method

All vorticity is represented by vortex lines that extend into the z direction. As these vortex lines originate on the vocal folds and are convected downstream, they form two continua that appear as curves in the x-y plane. These continua are vortex sheets in the limit of large Reynolds numbers. The two vortex sheets separate the glottal jet from the surrounding quiescent air in the rest of the duct.

The equation for Green's function, Eq. (7), can be used to simplify Eq. (10), the vortex sound term, in the limit $x \to +0$. In this limit the acoustic time delay is zero, and the time integral in Eq. (10) can be evaluated using an asymptotic approximation, so that the remaining spatial integral is evaluated at time t. Further, integrating in the z direction, Eq. (10) can be written in simplified form

$$B_\omega(\mathbf{x},t) = \frac{1}{4h}\int \left[\omega\left(-v_{y'}\frac{\partial Y}{\partial x'} + v_{x'}\frac{\partial Y}{\partial y'} \right) \right] (\mathbf{y},t))d^2\mathbf{y}, \tag{12}$$

where ω is the signed magnitude of vorticity, and $\mathbf{v}=v_{x'}\hat{\mathbf{i}}+v_{y'}\hat{\mathbf{j}}$ is the vorticity convection velocity.

In this two-dimensional geometry, the axial component of the Kirchhoff vector Y is the real part of a complex potential. The imaginary part is the corresponding axial Kirchhoff stream function ψ, say. Thus, Eq. (12) can be rewritten as

$$B_\omega = -\frac{1}{4h}\int [\omega(\mathbf{v}\cdot\nabla\psi)](\mathbf{y},t)d^2\mathbf{y}. \tag{13}$$

Let $\mathbf{y}_\omega(t)$ be the position of the line vortex ω in the x-y plane. $\mathbf{y}_\omega(t)=\mathbf{y}_s(\tau_w)+\int_{\tau_\omega}^{t}\mathbf{v}dt'$, and τ_w is the time that vortex ω is shed from the vocal fold surface at $\mathbf{y}=\mathbf{y}_s(\tau_\omega)$. Under the conditions of sufficiently low frequency, $\mathbf{y}_\omega$ describes two continuous, one-dimensional curves, within the region of the vocal and ventricular folds—one originating on the upper folds, $y_s'>0$, denoted $\Omega^+=\Omega^+(t)$, and the other from the lower folds, $y_s'<0$, denoted $\Omega^-=\Omega^-(t)$. Let $s_\omega=s_\omega(t)$ be the distance along either vortex sheet from the shedding point at time t, and $s_\omega^\perp=s_\omega^\perp(t)$ be the distance measured from the outward normal to either vortex sheet. In a vortex sheet model,

$$\omega(s_\omega(t),t) = \pm U_\sigma(\tau_\omega)\delta(s_\omega^\perp(t)), \tag{14}$$

where $U_\sigma(t)$ is the asymptotic jet velocity predicted by free streamline theory and $\delta(\cdot)$ is the one-dimensional Dirac delta function. The plus sign in Eq. (14) corresponds to the vorticity shed from the upper vocal fold with $y_s'>0$ and the minus sign to vorticity shed from the lower vocal fold with $y_s'<0$. The magnitude of the convection velocity of the vorticity between the jet and surrounding fluid depends on the strength of the vorticity in terms of jet velocity U_σ, and the direction of convection is approximately the axial direction (x-axis).

$$\mathbf{v}(\mathbf{y},t) = \tfrac{1}{2}U_\sigma(\tau_\omega)\hat{\mathbf{i}}. \tag{15}$$

Further, by symmetry, $\partial\psi/\partial x'((x',y'),t)=-\partial\psi/\partial x'((x',-y'),t)$. These considerations show that the contributions by each vortex sheet to the integral in Eq. (13) are equal, so that it is sufficient to consider the effect of the vortex sheet for $y_s'<0$ in the integral and to double the result. The particle velocity in the jet is related to the particle velocity in the duct by mass conservation

$$U_\sigma(t) = \left(\frac{2h}{\sigma\Delta_m(t)}\right)U(t), \tag{16}$$

where σ is the jet contraction ratio.

Combining Eqs. (14)–(16), Eq. (13) can be rewritten as

$$B_\omega(t) = \frac{h}{2\sigma^2}\int_{\Omega^-}\left(\frac{U(\tau_\omega)}{\Delta_m(\tau_\omega)}\right)^2\frac{\partial\psi}{\partial x'}ds_\omega. \tag{17}$$

It was shown in Howe and McGowan (2007) that the solution of the quasi-steady approximation to Eq. (9) [i.e., Eq. (9) with the time derivative term discarded] produced results very close to the solution of the full equation. Substituting from Eq. (17) into the quasi-steady version of Eq. (9) and solving for $U(t)$ produce

$$U(t) \approx \frac{p_I H(t)}{\rho_0 c_0}\left(1-\frac{\rho_0 h}{4p_I\sigma^2}\int_{\Omega^-}\left(\frac{U(\tau_\omega)}{\Delta_m(\tau_\omega)}\right)^2\frac{\partial\psi}{\partial x'}ds_\omega\right). \tag{18}$$

A time step of $\Delta t=0.000\,01$ s was used and the vortex sheet $\Omega^-(t)$ was approximated by discrete vortices shed at each time step. A fourth-order Runge–Kutta extrapolation was used to advance the positions of these discrete vortices. Then Eq. (18) was used to find the acoustic particle velocity at each time step, with a smoothing window that took a linear weighted average of the current value with the three previous values. The initial condition was $U(+0)=0$. The integral in Eq. (18) was approximated using the trapezoid rule, with the term corresponding to $\tau_\omega=t$ excluded, as it had a negligible effect. The stream function ψ was found at each time step by solving Laplace's equation numerically with uniform inflow and outflow conditions at $x=-2$ cm and $x=2$ cm. A standard second-order finite difference algorithm with a mesh of length 0.005 cm on each side was used for the potential flow solution.

The following were the values of the parameters appearing in this equation (all quantities are in c-g-s): $p_I=8000$, $\rho_0=0.0013$, $h=0.8$, $c_0=34\,800$, and $\ell_3=1$. The contraction ratio σ for vocal folds in high Reynolds number flow can be expected to be between 0.6 and 1.8 (Park and Mongeau, 2007), and its value was kept constant with the value unity. This is consistent with the assumption that the vorticity is convected in the x direction. Equation (11) was used to compute $\Delta_m(t)$ at each time step. The separation point $\mathbf{y}_s(t)$ was set to be the point on the surface of the vocal folds that was closest to the centerline of the duct, with one exception. The exception was when this entailed moving the separation point from the previous time step farther upstream than the distance that was traveled by the previously shed vortex in time Δt, in which case the separation point was set upstream from the previous point by that convection distance. Discontinuities in the vortex sheet were avoided by doing this. When the separation point moved downstream from the previous time step all the previously shed vortices that were upstream were presumed to reattach into the boundary layer.

Initially, numerical experiments were conducted at various fundamental frequencies f_0 and for various ventricular fold shapes. It became apparent that the effect of the ventricular folds on the outgoing acoustic particle velocity was small within the ranges of geometric parameters provided in Agarwal *et al.* (2003). The effect appeared to be largest when the ventricular folds were close together and the fundamental frequency was relatively high. Therefore, only one geometric condition with ventricular folds as previously described and shown in Fig. 2 will be examined below at relatively high f_0s of 325 and 650 Hz.

IV. RESULTS

The acoustic signal downstream to the source region was computed using Eq. (18). The results are presented in terms of transmitted pressure, normalized by the subglottal

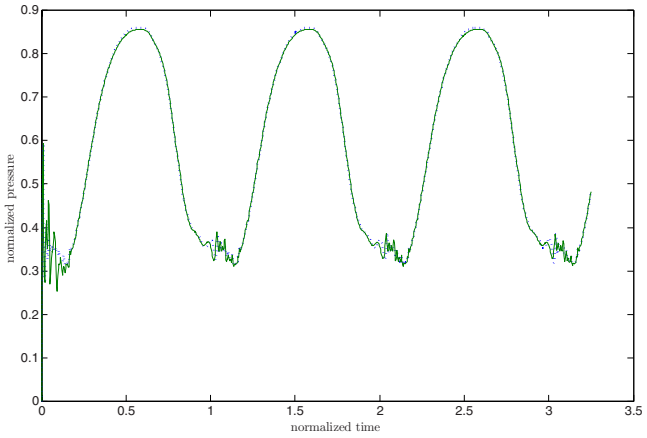


FIG. 3. (Color online) Propagated pressure wave downstream of the source region for when there are only vocal folds (dashed line) and when there are both vocal folds and ventricular folds (solid line) at $f_0=325$ Hz. Pressure is normalized by incident subglottal pressure p_I and time by $1/f_0$.

pressure in Figs. 3 and 4. Figure 3 compares the configuration with only vocal folds (dotted curve) to the configuration with ventricular folds added (solid curve) at $f_0=325$ Hz. The ventricular folds cause only a very slight decrease in peak transmitted pressure. Figure 4 shows the comparison at $f_0=650$ Hz. Again, the effect of the ventricular folds is very small.

Figure 5 shows the differences in the axial Kirchhoff streamlines for the case without ventricular folds [Fig. 5(a)] and the case with ventricular folds [Fig. 5(b)] when the vocal folds are at their most open position [one-half way through the cycle as defined by Eq. (10)]. The vortex sheets one-half through the first cycle are also indicated. Figure 6shows the gradient of ψ projected onto the vortex convection direction (the x direction), at the same phase of the glottal cycle shown in Fig. 5. Figure 6(a) is without ventricular folds and Fig. 6(b) is with ventricular folds.

Equation (18) involves integration along a vortex sheet, so it is important to be aware of its location and shape. Figure 7 shows the vortex sheet in steps of one-sixth of a cycle during the second cycle in the case of $f_0=325$ Hz with ventricular folds. Figure 8 shows the same sequence for $f_0=650$ Hz. The vortex sheet begins to bend over on itself in the first frame of the second glottal cycle in Fig. 7 and in the second frame in Fig. 8. At these times the glottis is in the process of opening, and the faster moving vortices that were shed when the glottis was the most constricted are catching up with the slower moving vortices that were shed when the glottis was the least constricted in the previous cycle. Because Figs. 7 and 8 have the abscissa in time normalized by the vocal fold vibration period, the bending of the vortex sheet occurs at about the same absolute time into the second cycle for both the 325 Hz case and the 650 Hz case. This can be expected because the vortex convection velocities follow very similar patterns. This follows by comparing Figs. 3 and 4 and relating vortex convection velocity to acoustic pressure using Eqs. (8), (14), and (15). The place where this bending occurs is farther upstream for the 650 Hz case because the slower moving vortices shed in the first cycle have not moved as far downstream in the second cycle as in the 325 Hz case.

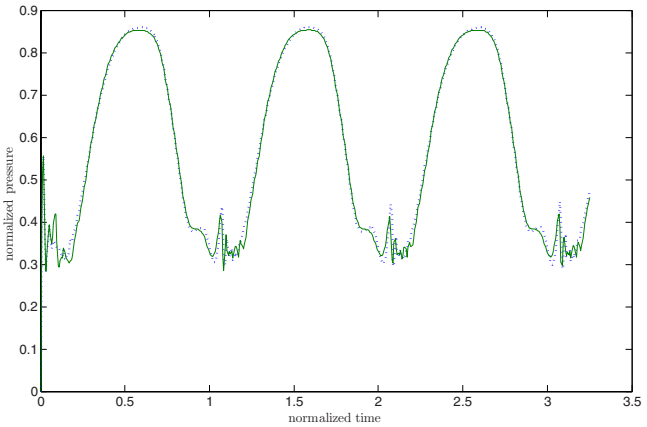


FIG. 4. (Color online) Propagated pressure wave downstream of the source region for when there are only vocal folds (dashed line) and when there are both vocal folds and ventricular folds (solid line) at $f_0=650$ Hz. Pressure is normalized by incident subglottal pressure p_I and time by $1/f_0$.

V. DISCUSSION

The numerical results showing little effect of the ventricular folds on the acoustic output are indicative of two general theorems that follow immediately from Eq. (13). First, in a steady flow, and for vorticity ultimately convected in the horizontal x direction, B_ω depends only on the difference between value of the ψ streamline on which the vorticity arises and the value of the "end" ψ streamline, which is the streamline on which the vortex sheet approaches as $x' \to +\infty$. This follows from the facts that, for a steady flow, both ω and $\mathbf{v}$ are constants in time along the vortex sheet, so that the integral in Eq. (13) has only the x-component of the ψ gradient in the integrand. A second theorem also holds under the same conditions of steady flow and ultimately horizontally convected vorticity: Changes in the shapes of the channel wall will not affect B_ω as long as the value of the streamline on which the vorticity arises and the value of where the vorticity "ends" are unchanged. In the numerical experiments performed in the present work, the streamlines from which vorticity arose (i.e., the streamline at the surface of the vocal fold) and the streamlines where the vorticity becomes parallel to the channel axis were the same with and without the presence of the ventricular folds. Note that these theorems do not depend on the exact path that the vorticity follows, only that the convection ultimately be in the x direction. So the assumption that the vorticity be convected axially is not critical to the conclusion that the ventricular folds do not have a large effect on acoustic output, at least if the vortex sheets do not strike the ventricular folds.

The flows here were not steady. However, these theorems remain approximately true for unsteady flow as long as vorticity traverses the regions of substantial nonzero x-component of the ψ gradient in a short time. Time of traversal is short if the strength of the vorticity entering the region of substantial nonzero x-component of the ψ gradient is nearly equal to the strength of the vorticity exiting this region. In the case of the vocal folds, this time should be short in relation to the inverse of the characteristic frequency

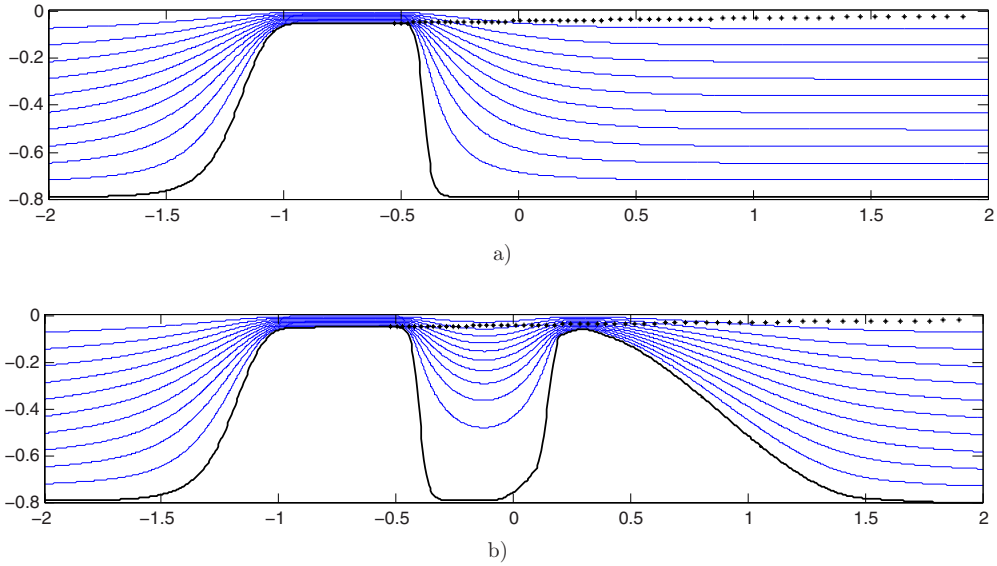


FIG. 5. (Color online) The streamlines of the horizontal component of the Kirchhoff vector and the vortex sheet at 1/2 of the vocal fold cycle, when the vocal folds are maximally open (a) for the case when there are only vocal folds and (b) for the case with ventricular folds.

of the variation in vortex strength. If d is the length scale of the region of substantial nonzero gradient, f_0 is a characteristic frequency for the variation of vortex strength, and V is the vorticity convection velocity in the same region, then the condition that $(f_0 d)/V \ll 1$ is necessary for each region of nonzero x-component of the ψ gradient. In the middle of the glottal cycle shown in Fig. 6(b), there is a region of substantial positive x-component of gradient at the trailing edge of the vocal folds, followed by a region of substantial negative x-component of gradient before the leading edge of the ventricular folds, and a region of substantial positive x-component of gradient after the peak of the ventricular folds. The length scale for these regions together is of the order of 1.0 cm. The vortex convection velocities range from about 1500 to 2600 cm/s, so at 325 and 650 Hz $(f_0 d)/V$ is at most 0.4, and often much smaller.

Another way of interpreting the weak effect on voicing of the ventricular folds is as follows. Vorticity shed from the

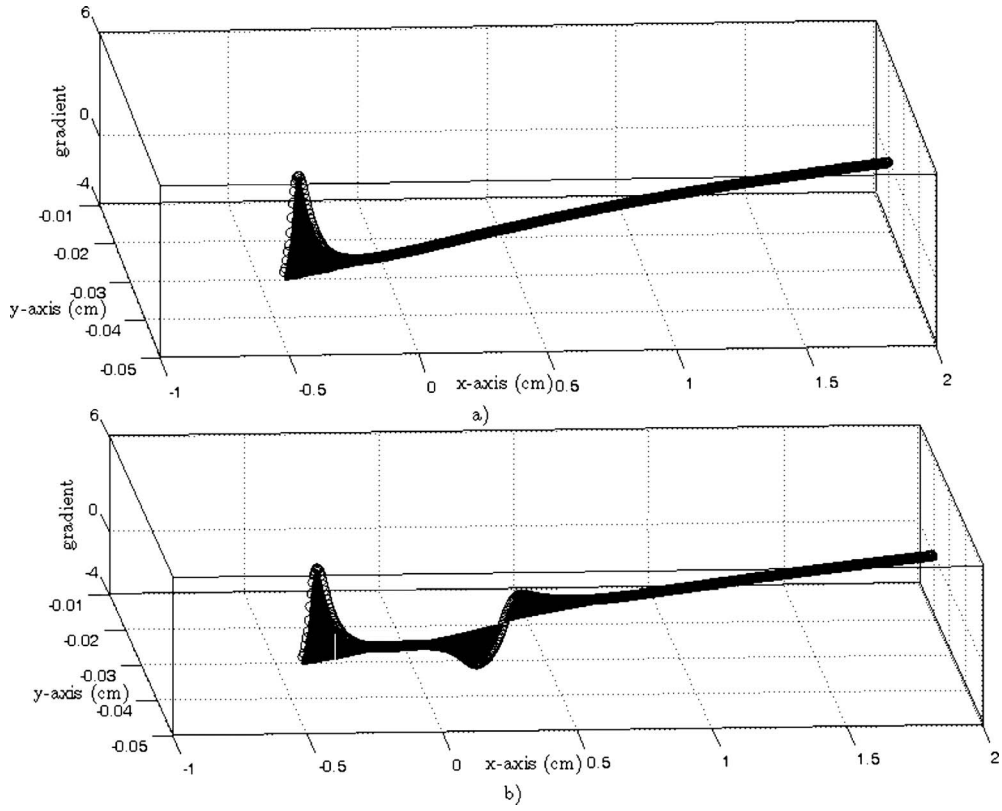


FIG. 6. Gradient of the axial Kirchhoff stream function projected onto the vortex convection direction (x-axis) as a function of position of the elements of the vortex sheet at 1/2 of the vocal fold cycle (a) for the case when there are only vocal folds and (b) for the case with ventricular folds.

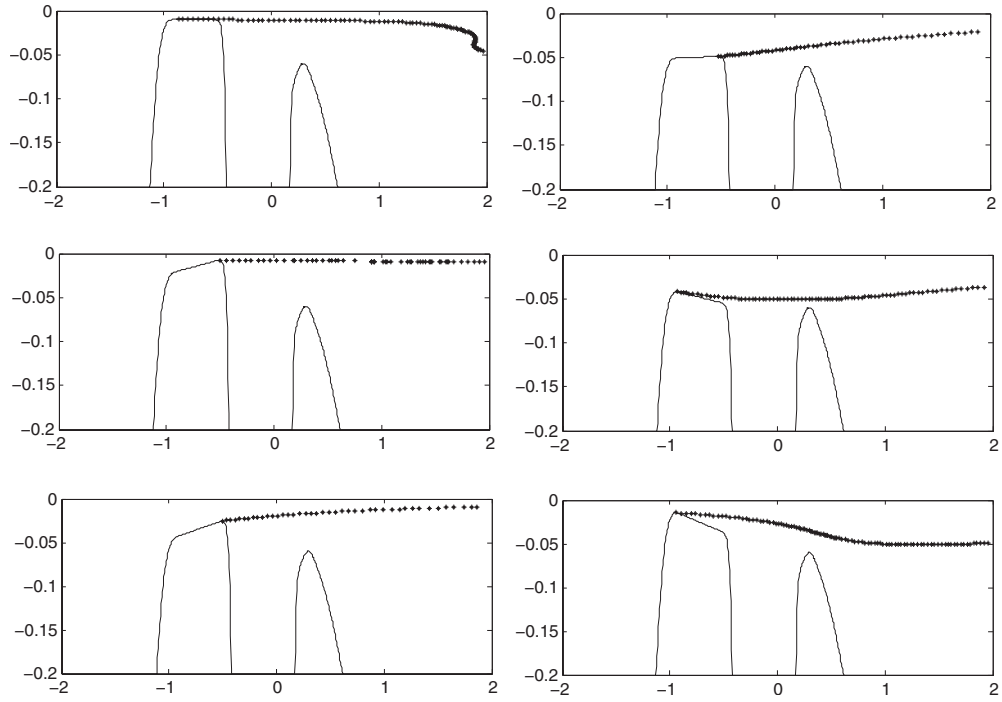


FIG. 7. Vortex sheet at $f_0=325$ Hz shown with outlines of the top of the vocal folds and the ventricular folds at 1/6 intervals of the second glottal cycle. The columns of panels on the left are during the opening phase, sequenced from top to bottom. The columns on the right are during the closing phase, sequenced from top to bottom. The y-axis has been magnified with respect to the x-axis.

vocal folds always passes through the positive x-component of ψ gradient region, creating the fluctuating drag force that is propagated as sound down the vocal tract. The ventricular folds also create their own nonzero regions of x-component of gradient, strongest near the top of the ventricular folds, but with a negative and a positive x-component of gradient regions that tend to cancel one another.

Figures 7 and 8 indicate that the vortex sheets will eventually fold onto themselves. The higher the fundamental frequency, the farther upstream this folding will occur, as indicated by comparing the two figures. The folding is a prelude to transitions to the vortex sheet breaking apart and, eventually, a transition to turbulence. At frequencies somewhat higher than 650 Hz examined here, the vortex sheet could

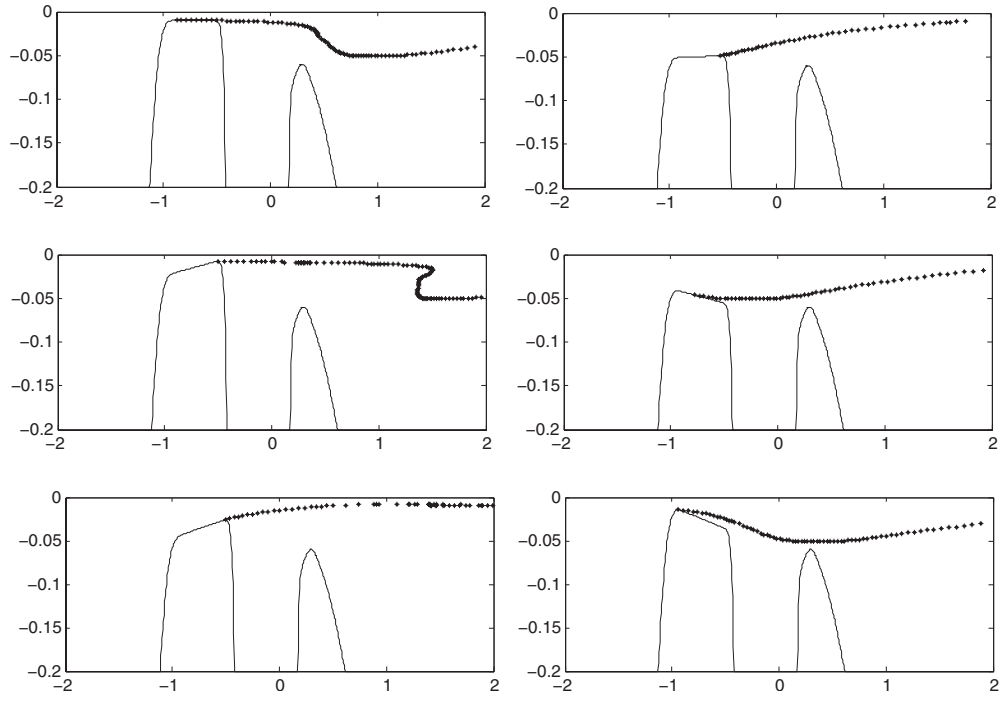


FIG. 8. Vortex sheet at $f_0=650$ Hz shown with outlines of the top of the vocal folds and the ventricular folds at 1/6 intervals of the second glottal cycle. The columns of panels on the left are during the opening phase, sequenced from top to bottom. The columns on the right are during the closing phase, sequenced from top to bottom. The y-axis has been magnified with respect to the x-axis.

break up before passing above the tip of the ventricular folds. In this case the ventricular folds could have a substantial effect on the output acoustics farther down the vocal tract. The vorticity would interact with the negative x-component of the ψ gradient upstream of the ventricular folds in a coherent manner, but not with the positive gradient downstream. This would promote flow from the glottis because it would act as a negative resistance to flow, thus increasing the peak flow through the glottis. Thus the model is restricted to frequencies for which the vortex sheet does not break up before interacting with both the upstream and downstream sides of the ventricular folds. Further there is some experimental evidence that this kind of breakup can occur for nominally steady jets. Agarwal (2004) experimented with glottal jets and rigid model ventricular folds and found that the presence of the ventricular folds decreased the resistance to flow when the tips of the false folds were separated by distances from one to three times greater than the minimum distance between the vocal folds. This would indicate that the jets remained laminar from the glottis to the channel formed by the tips of the model false folds and became turbulent downstream of this channel. However, the intensity of the inflow turbulence and the level of noise downstream of the test section were not quantified in these experiments.

When vortex sheets' breakup occurs the condition that $(f_0 d)/V$ be small no longer holds, and new effects could be expected to be observed. One such effect is a positive feedback of the vorticity at the ventricular folds enhancing the vortex shedding in the glottis and, hence, the glottal jet. This effect is known as Rossiter buffeting (Howe, 1998) at frequencies greater than 2 kHz. Zhang *et al.* (2002) did attribute the occurrence of a peak in their predicted acoustic spectrum to this phenomenon in their CFD experiment. Whether this requires the breakup of the vortex sheet as described above, or only a time delay in interaction between upstream and downstream portions of the ventricular folds is uncertain at this time. The breaking up of the vortex sheet as described above would certainly enhance the buffeting.

Vortex sheet modeling has been used here on the basis that the distance through a glottal jet shear layer is negligible compared to other length scales, such as glottal channel width. This is a good approximation for the initial portion of the glottal jet. However, the jet shear layer thickens with downstream distance, and, depending on the height of the ventricular folds, a part of the jet will strike the upstream portion of the ventricular folds. This would add to the drag force, which appears to have been the case for false fold tips closer together than were the vocal fold surfaces in Agarwal's (2004) experiments.

VI. CONCLUSION

The ventricular folds do not have a noticeable effect on the voice source according to these results when the movement of the vocal folds is specified. This confirms the CFD results of Zhang *et al.* (2002) at low frequencies. This conclusion is derived from aeroacoustic theory and the use of compact Green's functions so that several fundamental frequencies and geometrical configurations are easily tested. Most importantly, the numerical results can be confirmed by mathematical considerations regarding the integral of the gradient of the axial Kirchhoff stream function. This theoretical understanding is an important one for sources of sound in the vocal tract in general.

In a simulation where the vocal fold motion is not prescribed, but occurs because of interaction with the air, it can be argued that there still will be little effect of the ventricular folds on the oscillation at low frequencies. This is because the vorticity interacting with the negative portion of the gradient of the axial Kirchhoff stream function upstream of the ventricular folds is nearly equal in strength to the vorticity interacting with the positive gradient downstream of the ventricular folds. However, the effect of the ventricular folds on the movement of the vocal folds has not been analyzed here. Further, if the glottal jet strikes the ventricular folds, then their motion would need to be considered.

Similarly, the results presented here may be generalized to fully three-dimensional situations. All that is required is that the changes in vortex strength be small over the regions of nonzero gradients in the axial Kirchhoff stream function that the vorticity traverses.

ACKNOWLEDGMENT

This work was supported by a subaward of Grant No. DC-009299 from the National Institute on Deafness and Other Communication Disorders to the University of California, Los Angeles.

Agarwal, M. (**2004**). "The false vocal folds and their effect on translaryngeal airflow resistance," Ph.D. thesis, Bowling Green State University, Bowling Green, OH.

Agarwal, M., Scherer, R. C., and Hollien, H. (**2003**). "The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings," J. Voice **17**, 97–113.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands).

Howe, M. S. (**1975**). "Contributions to the theory of aerodynamic sound, with application to excess jet noise and the theory of the flute," J. Fluid Mech. **71**, 625–673.

Howe, M. S. (**1998**). *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge).

Howe, M. S. (**2003**). *Theory of Vortex Sound* (Cambridge University Press, Cambridge).

Howe, M. S., and McGowan, R. S. (**2007**). "Sound generation by aerodynamic sources near a deformable body, with application to voiced speech," J. Fluid Mech. **592**, 367–392.

Lighthill, M. J. (**1952**). "On sound generated aerodynamically. Part I: General theory," Proc. R. Soc. London, Ser. A **211**, 564–587.

Park, J. B., and Mongeau, L. (**2007**). "Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx," J. Acoust. Soc. Am. **121**, 442–455.

Rayleigh, L. (**1945**). *Theory of Sound* (Dover, New York), Vol. **2**.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Upper Saddle River, NJ).

van den Berg, Jw., Zantema, J. T., and Doornenbal, P. (**1957**). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 626–631.

Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

Biomechanical modeling of register transitions and the role of vocal tract resonators[a)]

Isao T. Tokuda[b)]
School of Information Science, Japan Advanced Institute of Science and Technology, Nomi-city, Ishikawa 923-1292, Japan

Marco Zemke
Institut für Informatik, Humboldt University Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

Malte Kob
Erich Thienhaus Institute, University of Music Detmold, Neustadt 22, 23756 Detmold, Germany

Hanspeter Herzel
Institute for Theoretical Biology, Humboldt University Berlin, Invalidenstraße 43, 10115 Berlin, Germany

(Received 3 March 2009; revised 27 July 2009; accepted 31 July 2009)

Biomechanical modeling and bifurcation theory are applied to study phonation onset and register transition. A four-mass body-cover model with a smooth geometry is introduced to reproduce characteristic features of chest and falsetto registers. Sub- and supraglottal resonances are modeled using a wave-reflection model. Simulations for increasing and decreasing subglottal pressure reveal that the phonation onset exhibits amplitude jumps and hysteresis referring to a subcritical Hopf bifurcation. The onset pressure is reduced due to vocal tract resonances. Hysteresis is observed also for the voice breaks at the chest-falsetto transition. Varying the length of the subglottal resonator has only minor effects on this register transition. Contrarily, supraglottal resonances have a strong effect on the pitch, at which the chest-falsetto transition is found. Experiment of glissando singing shows that the supraglottis has indeed an influence on the register transition.
 [DOI: 10.1121/1.3299201]

PACS number(s): 43.70.Bk, 43.70.Gr [DAB] Pages: 1528–1536

I. INTRODUCTION

Voice registers have been introduced for perceptually distinct types of certain vocal qualities that can be maintained over some ranges of pitch and loudness (Titze, 2000). The perceptive classification can be accompanied by measurements of voice source parameters such as spectral slope or glottal open quotient (Henrich *et al.*, 2005; Salomão and Sundberg, 2008). Characteristic features are vocal breaks at register transitions associated with pitch and amplitude jumps (Roubeau *et al.*, 1987; Švec *et al.*, 1999; Miller *et al.*, 2000). It has been shown by Hirano *et al.* (1970) that the thyroarytenoid (TA) muscle and the cricothyroid (CT) muscle regulate register transitions. Consequently, the perceptive aspects of registers are inherently related to laryngeal features of vocal fold vibrations. In chest phonation the vocal folds are thick and glottal closure is complete, whereas in falsetto only the vocal fold edges vibrate (Vilkman *et al.*, 1995). Despite extensive experimental investigations using acoustic signals (Hollien, 1974; Sundberg and Gauffin, 1979), electromyography (Shipp and McGlone, 1971), electroglottography (Henrich *et al.*, 2005), and videokymography (Švec *et al.*, 2008), many questions regarding register transitions remain open: what determines the pitch of involuntary register transition? What is the role of sub- and supraglottal resonances? Is there hysteresis at register transition?

In order to address these problems, biomechanical modeling can complement experimental studies. Even though some register-like phenomena have been described in two-mass models (Sciamarella and d'Alessandro, 2004; Zaccarelli *et al.*, 2006), an appropriate representation of vibratory modes in chest and falsetto requires more advanced models such as body-cover model of Story and Titze (1995), two-dimensional model of Adachi and Yu (2005), and three-mass model of Tokuda *et al.* (2007). Moreover, a smoothed glottal geometry improves the classical two-mass model of Ishizaka and Flanagan (1972) considerably (Pelorson *et al.*, 1994; Lous *et al.*, 1998).

In this paper, we use a four-mass body-cover polygon model (Tokuda *et al.*, 2008) to study register transitions and the influence of resonators. Sub- and supraglottal resonances are described using the wave-reflection model (Kelly and Lochbaum, 1962; Liljencrants, 1985; Story, 1995; Titze, 2006). Our model simulations reveal hysteresis at the phonation onset and at chest-falsetto transition, which is consistent with experimental data (Berry *et al.*, 1996; Horáček *et al.*, 2004). We find that vocal tract resonances have a pronounced effect on the chest-falsetto transition.

II. FOUR-MASS BODY-COVER POLYGON MODEL

There are complex high-dimensional models of vocal fold vibrations (Alipour *et al.*, 2000; Titze, 2006; Gömmel *et*

[a)]This paper is based on a talk presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: isao@jaist.ac.jp

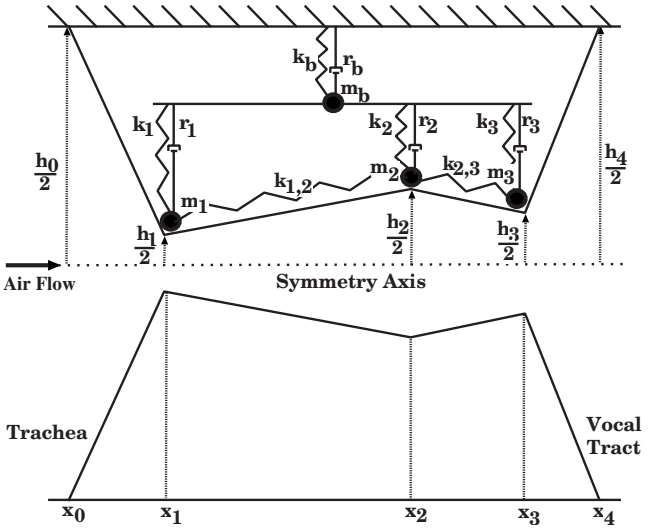


FIG. 1. Schematic illustration of the four-mass polygon model of the vocal folds.

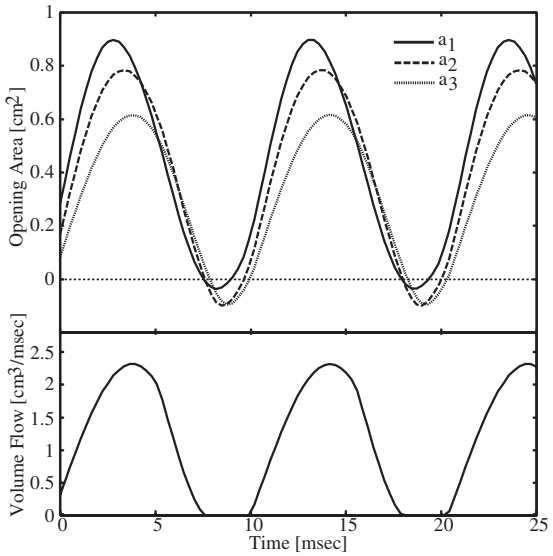


FIG. 2. Chest-like vibration simulated by the default parameter setting. Time series of glottal areas $\{a_i=lh_i:i=1,2,3\}$ (l is glottal length) between the three masses are shown in the upper graph, whereas the corresponding volume flow U is shown in the lower graph.

al., 2008; Zhang, 2009) describing anatomical and physiological details. However, many parameters are not precisely known and a comprehensive bifurcation analysis is difficult (Berry *et al.*, 1994). Consequently, we constructed a low-dimensional model that is consistent with the basic experimental observations.

Our model is based on the body-cover differentiation proposed by Story and Titze (1995), a three-mass representation of the cover (Tokuda *et al.*, 2007), and a smooth vocal fold geometry as in Lous *et al.* (1998). Advantage of dividing the cover layer into the three masses is that they are suitable for representing the coexistence of different vibratory patterns, which may correspond to chest and falsetto registers (Tokuda *et al.*, 2007). The chosen parameter values are similar to the values in these papers and they are consistent with muscle activation rules (Titze and Story, 2002). A detailed discussion of the modeling and a complete set of equations and parameters are given in Appendixes A and B and a recent thesis (Zemke, 2008).

Figure 1 visualizes our four-mass polygon model. The three cover masses allow wave-like vibrations of the whole vocal folds with a complete closure of the glottis in chest-register simulation. For other parameter sets, high-pitched oscillations with diminished closure of the glottis are simulated, which resemble falsetto register. Figure 2 shows chest-like vibrations (fundamental frequency of 96 Hz) for the default parameters listed in Appendix A. The phase shifts in the opening areas shown in the upper graph allow the energy transfer from the air flow to the masses and contribute to a skewing of the glottal pulses of the lower graph.

In order to simulate register transitions, we recall rules for controlling low-dimensional vocal fold models with muscle activation (Smith *et al.*, 1992; Titze and Story, 2002). An active CT muscle decreases the vibrating mass and increases stiffness. We introduce a tension parameter T which mimics the CT muscle (Steinecke and Herzel, 1995). Masses are divided by T and stiffness parameters are multiplied by T and thus the fundamental frequency of the model is roughly proportional to the parameter T. Increasing T from 1 to 7 allows pitch variation from 100 to 600 Hz exhibiting register transitions (detailed simulations in Sec. IV). Figure 3 shows an example of simulating falsetto-like vocal fold oscillations (fundamental frequency of 369 Hz) with the tension parameter $T=3.6$. An almost sinusoidal glottal volume flow with weaker harmonics is observed.

III. MODELING VOCAL TRACT RESONANCES

For male speech, the linear source-filter theory (Fant, 1960) was quite successful (Stevens, 1999). However, as shown already in Ishizaka and Flanagan (1972), source-filter coupling is essential if the fundamental frequency F_0 is comparable to the formant frequencies (Titze, 2008; Titze *et al.*, 2008). Register transition can be regarded as bifurcations of limit cycle oscillations (see Tokuda *et al.*, 2007 for a detailed

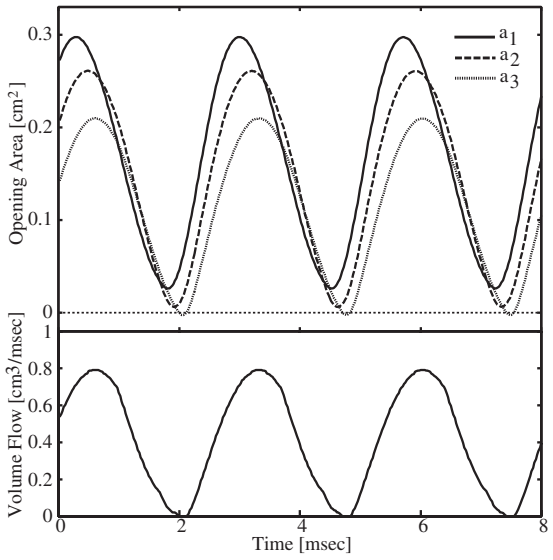


FIG. 3. Falsetto-like vibration simulated with the tension parameter of $T=3.6$.

discussion). It is known that bifurcations, i.e., sudden transitions due to slow parameter variations, might depend strongly on system parameters. Thus, it is possible that even medium effects of the resonances (see Titze *et al.*, 2008 for data on source-filter interactions) can shift register transition drastically. It has been suggested that particular subglottal resonances govern involuntary register transitions (Titze, 2000; Zhang *et al.*, 2006).

In order to study source-filter coupling, we implemented the wave-reflection model (Kelly and Lochbaum, 1962; Liljencrants, 1985; Story, 1995; Titze, 2006). Details of the simulations are given in Appendix B. For simplicity we approximate the resonators by uniform tubes characterized by their length and area. This simplification gives direct insight on how resonance frequencies given by the tube lengths affect the location of register transitions.

IV. SIMULATION RESULTS

A. Simulation of gliding pitch

The analysis of register transitions and source-tract interaction is often studied using glissando singing (Henrich *et al.*, 2005), experimental variation in vocal fold tension (Tokuda *et al.*, 2007), or gliding of the fundamental frequency in biomechanical models (Titze, 2008). In Fig. 4, we compare a glissando of an untrained singer with simulations of a corresponding F_0 glide in our four-mass model coupled to sub- and supraglottal resonators. The singer's glissando in Fig. 4(a) exhibits register transitions with frequency jumps around 3.3 and 7.8 s at slightly different pitches. There is an abrupt phonation onset at 1.2 s and a smoother offset with some irregularities. Glissando is simulated in Fig. 4(b) by varying our tension parameter T from 1 to 5.5 and then back. We find a frequency jump at 6.7 s ($T=3.8, F_0=390$ Hz) and a backward transition at 17 s ($T=3.3, F_0=350$ Hz). These differences between chest-falsetto and falsetto-chest transitions are a landmark of hysteresis (see Tokuda *et al.*, 2007 for a detailed discussion of bifurcations leading to hysteresis). Hysteresis indicates that there are coexisting vibratory regimes ("limit cycles") for a range of parameters. Moreover, hysteresis implies that there are voice breaks instead of *passagi* of trained singers.

In addition to register transitions, occasionally subharmonics are observed, e.g., at 8.7 and 14.1 s. It has been discussed earlier (Berry *et al.*, 1996; Tokuda *et al.*, 2007) that register transitions are often accompanied by nonlinear phenomena such as subharmonics and chaos. The gross features of the experimental and simulated F_0 glides in Fig. 4 are similar. The study of hysteresis at phonation onset/offset requires a more detailed Hopf bifurcation analysis.

B. Bifurcation diagrams of phonation onset

Phonation onset and offset can be studied in the context of Hopf bifurcations (see, e.g., Lucero, 1998; Mergell *et al.*, 2000). Hopf bifurcation theory describes the onset of self-sustained oscillations due to parameter variations. Smooth oscillation onset is associated with a supercritical Hopf bifurcation whereas hysteresis and amplitude jumps indicate a subcritical Hopf bifurcation (see, e.g., Guckenheimer and Holmes, 1983 for details).

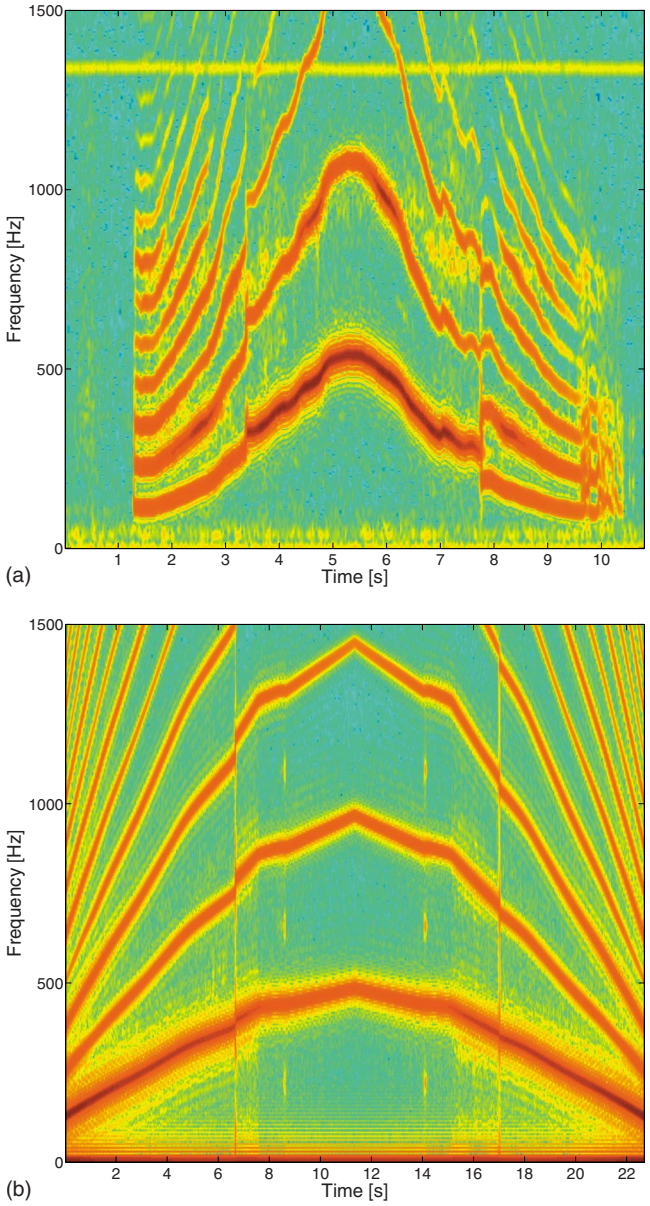


FIG. 4. (Color online) (a) Spectrogram of human voice with a gliding fundamental frequency (F_0). (b) Model simulation of the gliding F_0.

In case of phonation onset, an increasing subglottal pressure indicates vocal fold oscillations at overcritical values. In the simplified two-mass model (Steinecke and Herzel, 1995), no hysteresis was observed. In contrast, excised larynx experiments revealed a clear pressure difference between onset and offset values of about 0.2 kPa (Berry *et al.*, 1996).

Such a hysteretic phonation onset/offset is shown in our model simulations in Fig. 5. Figure 5(a) refers to the model without vocal tract resonators. There are amplitude jumps at 0.52 kPa (onset) and at 0.34 kPa (offset). Differences between increasing and decreasing pressures indicate a subcritical Hopf bifurcation. On the other side, Fig. 5(b) shows that sub- and supercritical bifurcations can occur in the four-mass model, to which sub- and supraglottal resonators are attached. As discussed in Titze (1988), the resonators reduce

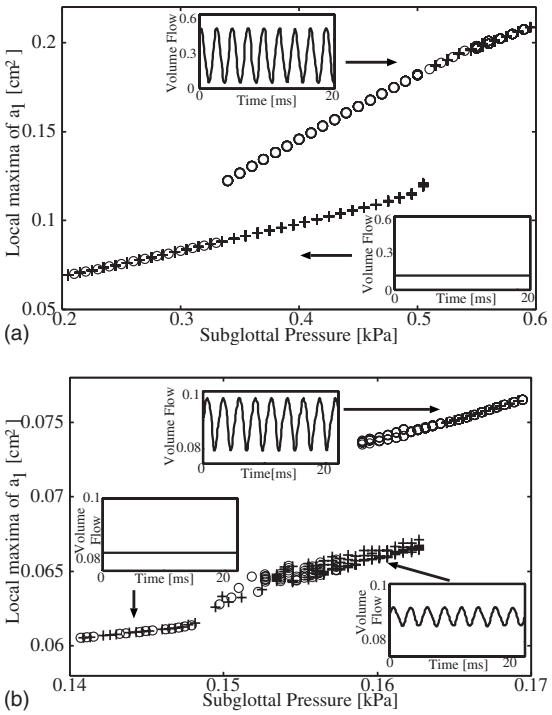


FIG. 5. Phonation onset. No vocal tract is attached to the vocal fold model in (a), but in (b) vocal tract is attached. Tension parameter is set as $T=4$. Local maxima of the opening area of the lower mass ($a_1=lh_1$) are plotted for both increasing (crosses) and decreasing (circles) subglottal pressure. The small graphs inside of each diagram represent the volume flow U (cm^3/ms) corresponding to each branch of the onset curve.

the threshold pressure. The phonation onset around 0.15 kPa seems rather smooth, implying a supercritical Hopf bifurcation. The generated phonation on the middle branch, however, has a relatively small amplitude and becomes unstable around 0.165 kPa, where it jumps to more stable one with larger amplitude on the upper branch. This jump creates hysteresis in the model simulations.

C. Influence of vocal tract on chest-falsetto transition

We induced register transitions in our model by changing the tension parameter T gradually and by measuring the fundamental frequency F_0, the amplitude of the opening area, and the number of the colliding cover masses. We observed a steady increase in F_0 and collision of all three cover masses at low F_0 and collision of only the top mass at high F_0. For simplicity, a binary classification is applied to draw the register transitions of Figs. 6–8 as follows: collision of three cover masses are termed chest, whereas collision of less masses are termed falsetto. If only the upper masses collide, open quotient (OQ), defined as OQ=duration of the open phase of the glottis/pitch period, became large as known from measurement in singers (Henrich *et al.*, 2005). In our bifurcation diagrams, with the tension T as the bifurcation parameter, we plotted the fundamental frequency F_0 on the x-axis instead of T since this allows a direct comparison with glissando spectrograms.

Figures 6 and 7 compare register transitions of the isolated four-mass model with the ones of the complete model including sub- and supraglottal resonances. In both cases, we

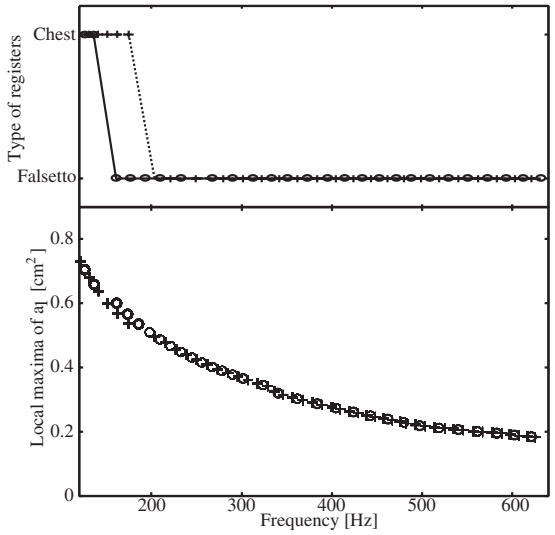


FIG. 6. Register transition of the vocal fold model without vocal tract. Frequency domains for chest and falsetto registers are drawn in the upper graph, whereas the corresponding bifurcation diagrams are drawn in the lower graph. The curves were drawn by both increasing (dotted line with crosses) and decreasing (solid line with circles) tension parameter T. In the bifurcation diagram, local maxima of the opening area of the lower mass ($a_1=lh_1$) were plotted.

find a pronounced hysteresis of about 30–40 Hz but relatively small jumps of the amplitudes. Most notable is the dramatic shift in the transition due to the coupling to vocal tract resonators. This observation reveals that the chest-falsetto transition depends sensitively on source-tract interactions.

In order to substantiate this finding, we varied the length of the sub- and supraglottal tubes. First, we decreased and increased the length of the subglottal tube by 25%. It turned out that there are only minor effects on the register transition. The length changes led to shifts in the transition point by 10–15 Hz (no graphs shown). In contrast, the supraglottal resonance had a profound effect: changing the default length of 17.5 cm to 75% or 125% induced major shifts in the

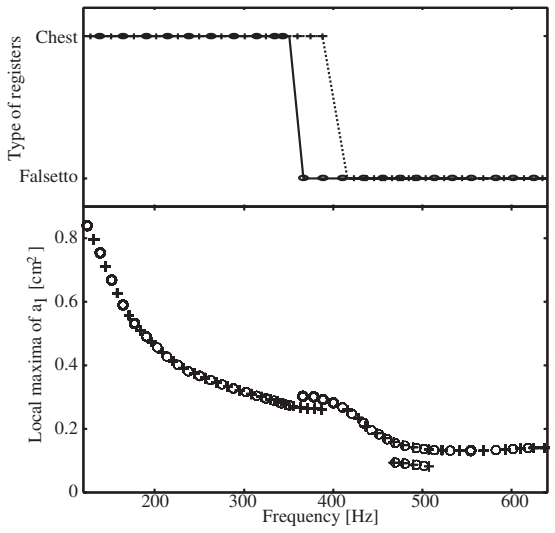


FIG. 7. Register transition of the vocal fold model with vocal tract. The default lengths for sub- and supraglottis are $L_{sub}=24.7$ cm and $L_{sup}=17.5$ cm, respectively.

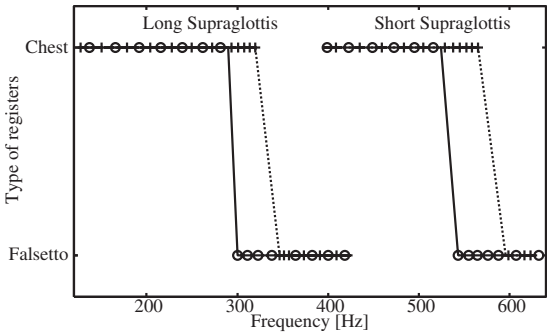


FIG. 8. Dependence of the register transition on the supraglottal length. The right and left graphs show the case of short supraglottis (L_{sup}=13.125 cm) and long supraglottis (L_{sup}=21.875 cm), respectively.

frequencies at which register transitions are observed (see Fig. 8). Note that these vocal tract lengths are within the physiological range. For all considered vocal tract length, subharmonics were observed slightly above the chest-falsetto transition. Hysteresis with a frequency difference of about 30–40 Hz and subharmonics were robust features of our simulation. The pitch of the register transition was, however, strongly affected by the formant frequencies.

V. EXPERIMENT

Our numerical study has shown that the supraglottal resonance has a strong influence on the pitch of the chest-falsetto transition. In order to examine this effect, we have carried out an experimental study of glissando on vowel /i/. This vowel has been chosen, since /i/ is one of the vowels that provide the lowest formant frequency F_1, thereby the F_0-F_1 interaction can be easily observed in the singing experiment. Four subjects were asked to perform F_0 gliding (from low F_0 to high F_0 and then back to low F_0). Two recordings were obtained from each subject. The subjects were all untrained males who have no evidence on laryngeal pathology. Both speech signal and electroglottographic (EGG) signal were simultaneously recorded.

Figure 9 shows an example of the recording data. The first formant F_1 was estimated from the speech signal by the conventional technique based on linear prediction analysis (McCandless, 1974). The spectrogram shows that, as the fundamental frequency F_0 increases and crosses the first formant F_1=250 Hz, the frequency jump is induced at t=4.6 s. The same frequency jump is observed, when the fundamental frequency decreases and crosses the first formant at t=6.5 s. As indicated by the OQ computed from the EGG signal by the method of Henrich *et al.* (2005), these frequency jumps are accompanied by the register change. The regime of falsetto register, characterized by high OQ, is clearly distinguished from the regime of chest register by dashed lines in Fig. 9(b). The timing of the register change coincides with the F_0-F_1 crossings quite well. This implies that the register transition is induced by the source-filter interaction, which is known to become strong when F_0 and F_1 are close to each other (Story *et al.*, 2000; Titze, 2008). Among the four subjects, coincidence of the register transition and the F_0-F_1 crossing has been observed in three subjects, where the other subject showed register transitions with a pitch much higher than the

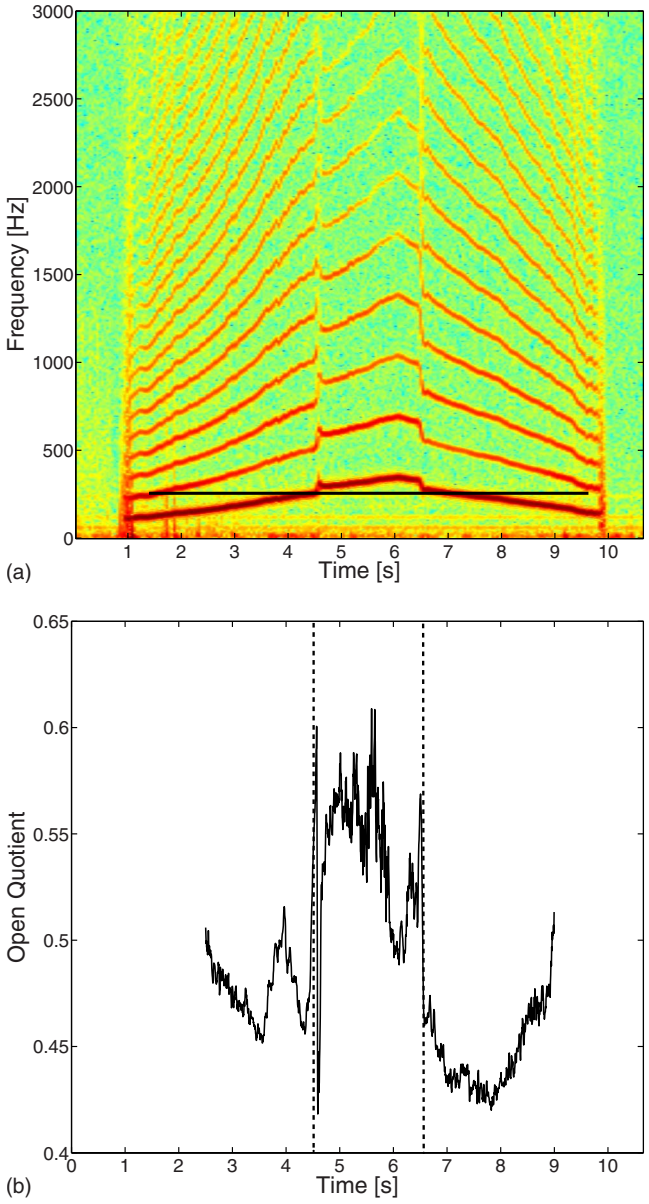


FIG. 9. (Color online) (a) Spectrogram of a male subject on vowel /i/ with a gliding fundamental frequency (F_0). The first formant (F_1) estimated from the speech signal is indicated by a solid line. (b) OQ computed from the EGG signal of (a). A region of high OQ, corresponding to falsetto register, is separated from low OQ regions by dashed lines.

first formant. Our experimental study therefore implies that the supraglottal resonance has indeed a strong influence on the register transitions. This influence of course depends on the individual characteristics of the subject, where well-trained singers should know how to sustain the chest register by avoiding the voice instability. Hence it is reasonable that not all subjects in our experiment showed a clear influence of the resonator on the registers.

To study how the different vowels affect the register transition, we further collected statistical data from the three subjects, who showed the influence of the resonator. Each subject was asked to perform F_0 gliding on both vowels /a/ and /i/, where ten recordings were obtained for each vowel. Average and standard deviations of the fundamental frequency F_0, at which the register transition takes place when

TABLE I. Average and standard deviation of the fundamental frequency F_0, at which the register transition takes place when increasing F_0. For each subject, ten recordings were collected.

Subject	I (Hz)	II (Hz)	III (Hz)
/a/	270 ± 19	271 ± 18	238 ± 20
/i/	240 ± 14	275 ± 13	231 ± 20

increasing F_0, were computed from the ten data sets, as summarized in Table I. Because of the high variability of the register transitions, the standard deviation was estimated to be relatively large. According to Welch's t-test, the mean frequency difference between /a/ and /i/ was statistically significant for subject I with a level of 1%. For the other two subjects, the difference was not significant.

We remark that Titze *et al.* (2008) carried out the same experimental framework in the context of voice instability induced by the source-tract coupling. Their main focus was, however, on the frequency jumps and not much attention has been paid to the register change. They found many frequency jumps induced by the F_0-F_1 crossing accompanied by hysteresis, in particular, for male subjects. Our observation essentially agrees with their study.

VI. SUMMARY AND DISCUSSION

From the nonlinear dynamics point of view, voice registers are distinct types of limit cycle oscillations. In this context, phonation onset refers to a Hopf bifurcation and register transitions are associated with bifurcations of limit cycles. In Tokuda *et al.* (2007), we characterized register transition in excised larynx experiments and simulations by two-dimensional bifurcation diagrams. In that paper, we analyzed a simple three-mass cover model.

Here we introduced a more realistic four-mass polygon model coupled to sub- and supraglottal resonators. This model was used to study bifurcations at phonation onset and the chest-falsetto transition.

In Mergell *et al.* (2000), a smooth phonation onset was quantified using the normal form of a supercritical Hopf bifurcation. This model explained high speed glottographic data in a reasonable way. In excised larynx experiments, however, amplitude jumps and hysteresis were reported at phonation onset (Berry *et al.*, 1996). The present simulation without vocal tract exhibits a subcritical Hopf bifurcation, where the associated hysteresis is in good agreement with the excised larynx experiments. Another simulation with vocal tract showed that the coupling to the resonators lowers the phonation onset threshold. It remains to be tested experimentally under which circumstances the phonation onset can be regarded as super- or subcritical Hopf bifurcation.

It is well known that register transitions in untrained singers are accompanied by vocal breaks (see, e.g., Švec and Pešák, 1994). In our simulations, we find indeed sudden jumps of pitch and amplitudes while varying the tension parameter T smoothly. In computer simulations, the associated phenomenon of hysteresis can be studied more easily than in glissando of singers. It turns out that our model exhibits clear pitch differences between chest-falsetto and falsetto-chest transitions.

In order to analyze the role of sub- and supraglottal resonators on register transition, we varied the length of the tubes by $\pm 25\%$. We find that the length of the subglottal tube has only minor effects on the register transition. Contrarily, the supraglottal resonator influences the pitch of the chest-falsetto transition strongly. To examine this effect, a simple experiment has been carried out based on the glissando singing on vowels /i/ and /a/. We have found a strong correlation between the register transition and the source-filter interaction for most of the subjects. Significant difference in the register transition point between the sung vowels was also detected from one subject despite high variability of the register breaks. These results provide a good indication that the supraglottal resonator has indeed an influence on the register transitions.

We remark that the present experiment is only preliminary and further investigations with more recording trials and with more subjects having various singing backgrounds are indispensable. To further study the influence of the resonators, the following experiments may also be of great interest. (i) The vocal tract anatomy can affect the register shift, i.e., vocal tract length might be correlated with the pitch of register transitions. (ii) Singing into tube (Hatzikirou *et al.*, 2006) should shift register transitions.

We finally note that comparable frequency jumps are also frequent in a variety of animal vocalizations (Wilden *et al.*, 1998; Tembrock, 1996; Riede and Zuberbühler, 2003).

ACKNOWLEDGMENTS

We thank Markus Hess and Frank Müller for the opportunity of acoustic and EGG recordings. We are grateful to Tobias Riede for stimulating discussions. This work was supported by SCOPE (Grant No. 071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

APPENDIX A: DETAILED FOUR-MASS MODEL

Figure 1 shows a schematic illustration of the four-mass polygon model. Following the body-covered theory (Hirano, 1974; Hirano and Kakita, 1985) this model is composed of a body part m_b and a cover part, which is divided into three masses m_i (lower: $i=1$, middle: $i=2$, and upper: $i=3$). Following the simplifications of (1) neglecting the cubic nonlinearities of the oscillators, (2) neglecting the additional pressure drop at inlet and considering the Bernoulli flow only below the narrowest part of the glottis (Steinecke and Herzel, 1995), and (3) assuming symmetry between the left and the right vocal folds, the model equations read as

$$m_1\ddot{y}_1 + r_1(\dot{y}_1 - \dot{y}_b) + k_1(y_1 - y_b) + \Theta(-h_1)c_1\left(\frac{h_1}{2}\right) + k_{1,2}(y_1 - y_2) = F_1, \tag{A1}$$

$$m_2\ddot{y}_2 + r_2(\dot{y}_2 - \dot{y}_b) + k_2(y_2 - y_b) + \Theta(-h_2)c_2\left(\frac{h_2}{2}\right) + k_{1,2}(y_2 - y_1) + k_{2,3}(y_2 - y_3) = F_2, \tag{A2}$$

$$m_3\ddot{y}_3 + r_3(\dot{y}_3 - \dot{y}_b) + k_3(y_3 - y_b) + \Theta(-h_3)c_3\left(\frac{h_3}{2}\right) + k_{2,3}(y_3 - y_2) = F_3, \tag{A3}$$

$$m_b\ddot{y}_b + r_b\dot{y}_b + k_b y_b + r_1(\dot{y}_b - \dot{y}_1) + k_1(y_b - y_1) + r_2(\dot{y}_b - \dot{y}_2) + k_2(y_b - y_2) + r_3(\dot{y}_b - \dot{y}_3) + k_3(y_b - y_3) = 0. \tag{A4}$$

The dynamical variables y_i represent displacements of the masses m_i, where the corresponding glottal opening is given by $h_i = h_{0i} + 2y_i$ (h_{0i} is prephonatory length; $i = 1,2,3$). The constant parameters r_i, k_i, and c_i represent damping, stiffness, and collision stiffness of the masses m_i, respectively, whereas $k_{i,j}$ represents coupling strength between two masses m_i and m_j. The stiffness is determined as $r_i = 2\zeta_i\sqrt{m_i k_i}$ using the damping ratio ζ_i. The collision function is approximated as $\Theta(\xi) = 0$ ($\xi \le 0$); $\Theta(\xi) = 1$ ($0 < \xi$).

The aerodynamic force, F_i, acting on each mass is derived as follows. First, the vocal fold geometry is described by a pair of four-mass-less plates, as shown in Fig. 1. The flow channel height $h(x,t)$ is a piecewise linear function, composed of $h_{1,0}$ ($x_0 \le x \le x_1$), $h_{2,1}$ ($x_1 \le x \le x_2$), $h_{3,2}$ ($x_2 \le x \le x_3$), and $h_{4,3}$ ($x_3 \le x \le x_4$), which are determined as

$$h_{i,i-1}(x,t) = \frac{h_i(t) - h_{i-1}(t)}{x_i - x_{i-1}}(x - x_{i-1}) + h_{i-1}(t), \tag{A5}$$

where $i = 1,2,3,4$ and h_0 and h_4 are constants. Assuming Bernoulli flow, the pressure distribution $P(x,t)$ below the narrowest part of the glottis, $h_{\min} = \min(h_1, h_2, h_3)$, is described as

$$P_s = P(x,t) + \frac{\varrho}{2}\left(\frac{U}{h(x,t)l}\right)^2 = P_0 + \frac{\varrho}{2}\left(\frac{U}{h_{\min}l}\right)^2, \tag{A6}$$

where ϱ represents the air density ($\varrho = 1.13$ kg/m^3), P_s is the subglottal pressure, P_0 is the supraglottal pressure, and l is the length of the glottis. In the case that no resonators are attached to the vocal folds, the subglottal pressure P_s is considered to be constant ($P_s = 0.8$ kPa), whereas the supraglottal pressure is assumed zero ($P_0 = 0$ kPa). This gives a simple formula for the glottal volume flow velocity as $U = \sqrt{2P_s/\varrho}h_{\min}l\Theta(h_{\min})$. The aerodynamic forces on the plates are induced by the pressure $P(x,t)$ along the flow channel. As m_1, m_2, and m_3 support the plates, an aerodynamic force on point i ($i = 1,2,3$) is found to be

$$F_i(t) = \int_{x_{i-1}}^{x_i} l\frac{x - x_{i-1}}{x_i - x_{i-1}}P(x,t)dx + \int_{x_i}^{x_{i+1}} l\frac{x_{i+1} - x}{x_{i+1} - x_{i-1}}P(x,t)dx. \tag{A7}$$

As shown by Lous *et al.* (1998), this integral can be solved analytically for the pressure distribution $P(x,t)$ described by Eq. (A6).

TABLE II. Default parameters of the four-mass polygon model.

Parameter	Symbol	Nominal value
Cover mass 1	m_1	0.009 g
Cover mass 2	m_2	0.009 g
Cover mass 3	m_3	0.003 g
Body mass	m_b	0.05 g
Stiffness of cover mass 1	k_1	6.0 N/m
Stiffness of cover mass 2	k_2	6.0 N/m
Stiffness of cover mass 3	k_3	2.0 N/m
Stiffness of body mass	k_b	30.0 N/m
Stiffness between cover masses 1 and 2	$k_{1,2}$	1.0 N/m
Stiffness between cover masses 2 and 3	$k_{2,3}$	0.5 N/m
Damping ratio of cover mass 1	ζ_1	0.1
Damping ratio of cover mass 2	ζ_2	0.4
Damping ratio of cover mass 3	ζ_3	0.4
Damping ratio of body mass	ζ_b	0.4
Collision stiffness of cover mass 1	c_1	$3k_1$
Collision stiffness of cover mass 2	c_2	$3k_2$
Collision stiffness of cover mass 3	c_3	$3k_3$
Prephonatory displacement of cover mass 1	h_{01}	0.036 cm
Prephonatory displacement of cover mass 2	h_{02}	0.036 cm
Prephonatory displacement of cover mass 3	h_{03}	0.036 cm
Height at vocal fold entrance	h_0	1.8 cm
Height at vocal fold exit	h_4	1.8 cm
Mass displacement at point 0	x_0	0 cm
Mass displacement at point 1	x_1	0.05 cm
Mass displacement at point 2	x_2	0.2 cm
Mass displacement at point 3	x_3	0.275 cm
Mass displacement at point 4	x_4	0.3 cm
Glottal length	l	1.4 cm

Parameter values used as the default situation of the present study are summarized in Table II. These values have been carefully selected in accordance with the previous studies (Ishizaka and Flanagan, 1972; Story and Titze, 1995; Lous *et al.*, 1998; Titze and Story, 2002). The observed phenomena were in general robust and did not show a strong dependence on the selected parameter values. To simulate the coexistence of the chest and falsetto registers, the damping ratio of the lower mass was set to be relatively small compared with the other masses. The small damping ratio activates the lower mass so that it leads to a large movement of all the cover masses to produce a chest-like register, which can easily coexist with a falsetto-like register.

The tension parameter T is also introduced to control the frequency of the four masses as

$$m_i' = m_i/T, \tag{A8}$$

$$k_i' = k_i \cdot T \quad (i = 1,2,3,b). \tag{A9}$$

The initial values for all simulations were set as $x_1 = 0.02$ cm, $x_2 = 0.015$ cm, $x_3 = 0.01$ cm, $x_b = 0$ cm, and $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = \dot{x}_b = 0$ cm/s. To integrate the four-mass model equations (A1)–(A4), Euler's method was applied with an integration step of $\Delta t = 11.4/8$ μs. The model equations were simulated also by using the MATLAB ODE solver

(ODE45) and it was confirmed that essentially the same results can be obtained.

To draw the bifurcation diagrams of Figs. 5–7, 20 local maxima of the opening area of the lower mass ($a_1=lh_1$) were plotted after discarding the transients. For the next parameter values, the final state of the preceding simulation was used as the initial condition.

The spectrogram of Fig. 4(b) was computed using the minimum glottal area ($a_{\min}=lh_{\min}$) with the following parameters. Sampling rate of 44 kHz, window length of 8192 sample points, overlap of 496 sample points, and Hanning window.

APPENDIX B: WAVE-REFLECTION MODEL FOR SUB- AND SUPRAGLOTTIS

Sub- and supraglottal resonances were described by using the wave-reflection model (Kelly and Lochbaum, 1962; Liljencrants, 1985; Story, 1995; Titze, 2006), which is a time-domain model of the propagation of one-dimensional planar acoustic waves through a collection of uniform cylindrical tubes. The supraglottal system was modeled as a simple uniform tube (area of 3 cm^2 and length of 17.5 cm), which is divided into 44 cylindrical sections. The area function for the subglottal tract was based on the one proposed by Zañartu *et al.* (2007). The area function is composed of 62 cylindrical sections. For both sub- and supraglottal systems, the section length Δz was set to 17.5/44 cm. This determines the sampling time interval as $\Delta t=\Delta z/c=11.4$ μs, where $c=350$ m/s stands for the sound velocity. The corresponding sampling frequency is 88 kHz.

Attenuation factor for the resonators was approximated as $a_k=1-0.007(\pi/A_k)^{1/2}\Delta z$ (A_k is kth cylinder area). Radiation resistance and radiation inertance at the lip were

$$R_r=\frac{128}{9\pi^2}\frac{\rho c}{A_L},\quad I_r=\frac{8}{3\pi^{3/2}}\frac{\rho}{\sqrt{A_L}}, \tag{B1}$$

respectively, where the lip area A_L was set to be equal to the last section of the supraglottis.

To couple the sub- and supraglottal resonators to the vocal fold model, an interactive source-filter coupling was realized according to Titze (2006, 2008). In this formula, the glottal flow is given by

$$U=\frac{a_g}{k_t}\left\{-\left(\frac{a_g}{A^*}\right)\pm\left[\left(\frac{a_g}{A^*}\right)^2+\frac{2k_t}{\rho c^2}(P_l+2p_s^+-2p_e^-)\right]^{1/2}\right\}, \tag{B2}$$

where $A^*=A_sA_e/(A_s+A_e)$ with A^s and A^e being the subglottal and supraglottal entry areas, respectively. k_t is a transglottal pressure coefficient set as 1.0. P_l stands for the lung pressure, whereas p_s^+ and p_e^- represent the incident partial wave pressures arriving from the subglottis and supraglottis, respectively. In the present study, subglottal and supraglottal entry areas were set to be equal to that of the last section of the subglottal system and that of the initial section of the supraglottal system, respectively. The lung pressure was set as $P_l=1.2$ kPa. Since subglottal pressure P_s is time dependent in this formula, the pressure value was averaged over a long-term simulation to plot P_s in Fig. 5(b).

Adachi, S., and Yu, J. (**2005**). "Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing," J. Acoust. Soc. Am. **117**, 3213–3224.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillation with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). "Bifurcations in excised larynx experiments," J. Voice **10**, 129–38.

Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Moulton, The Hague, The Netherlands).

Gömmel, A., Butenweg, C., and Kob, M. (**2008**). "Calculation model of the influence of the vocal fold shape and the ventricular folds on the laryngeal flow," in Proceedings of the 6th International Conference on Voice Physiology and Biomechanics, Tampere, Finland, pp. 194–196.

Guckenheimer, J., and Holmes, P. (**1983**). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer-Verlag, New York).

Hatzikirou, H., Fitch, W. T., and Herzel, H. (**2006**). "Voice instabilities due to source-tract interactions," Acta Acust. **92**, 468–475.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2005**). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J. Acoust. Soc. Am. **117**, 1417–1430.

Hirano, M. (**1974**). "Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniatr (Basel) **26**, 89–94.

Hirano, M., and Kakita, Y. (**1985**). "Cover-body theory of vocal cord vibration," in *Speech Science*, edited by R. G. Daniloff (College Hill Press, San Diego, CA), pp. 1–46.

Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch, and intensity of voice," Folia Phoniatr (Basel) **22**, 1–20.

Hollien, H. (**1974**). "On vocal registers," J. Phonetics **2**, 125–143.

Horáček, J., Švec, J. G., Veselý, J., and Vilkman, E. (**2004**). "Bifurcations in excised larynges caused by vocal fold elongation," in *Proceedings of the International Conference on Voice Physiology and Biomechanics*, edited by A. Giovanni, P. Dejonckere, and M. Ouaknine, Laboratory of Audio-Phonology, Marseille, France, pp. 87–89.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Kelly, J. L., and Lochbaum, C. (**1962**). "Speech synthesis," in Proceedings of the 4th International Congress on Acoustics, Paper No. G42, pp. 1–4.

Liljencrants, J. (**1985**). "Speech synthesis with a reflection-type line analog," Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden.

Lous, N. J., Hofmans, G. C., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two mass vocal fold model coupled to vocal tract and trachea, with application to prothesis design," Acta Acust. **84**, 1135–1150.

Lucero, J. C. (**1998**). "Subcritical Hopf bifurcation at phonation onset," J. Sound Vib. **218**, 344–349.

McCandless, S. (**1974**). "An algorithm for automatic formant extraction using linear prediction spectra," IEEE Trans. Acoust., Speech, Signal Process. **22**, 135–141.

Mergell, P., Herzel, H., and Titze, I. R. (**2000**). "Irregular vocal fold vibration—High-speed observation and modeling," J. Acoust. Soc. Am. **108**, 2996–3002.

Miller, D. G., Švec, J. G., and Schutte, H. K. (**2002**). "Measurement of characteristic leap interval between chest and falsetto registers," J. Voice **16**, 8–19.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation," J. Acoust. Soc. Am. **96**, 3416–3431.

Riede, T., and Zuberbühler, K. (**2003**). "Pulse register phonation in Diana monkey alarm calls," J. Acoust. Soc. Am. **113**, 2919–2926.

Roubeau, B., Chevrie-Muller, C., and Arabia-Guidet, C. (**1987**). "Electroglottographic study of the changes of voice registers," Folia Phoniatr (Basel) **39**, 280–289.

Salomão, G. L., and Sundberg, J. (**2008**). "Relation between perceived voice register and flow glottogram parameters in males," J. Acoust. Soc. Am. **124**, 546–551.

Sciamarella, D., and d'Alessandro, C. (**2004**). "On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters," Acta Acust. **90**, 746–761.

Shipp, T., Robert, E., and McGlone, R. E. (**1971**). "Laryngeal dynamics

associated with voice frequency change," J. Speech Hear. Res. **14**, 761–768.
Smith, M. E., Berke, G. S., Gerrat, B. R., and Kreiman, J. (**1992**). "Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration," J. Speech Hear. Res. **35**, 545–554.
Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal fold model," J. Acoust. Soc. Am. **97**, 1874–1884.
Stevens, K. (**1999**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Story, B. H. (**1995**). "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. thesis, University of Iowa, Iowa City, IA.
Story, B. H., Laukkanen, A.-M., and Titze, I. R. (**2000**). "Acoustic impedance of an artificially lengthened and constricted vocal tract," J. Voice **14**, 455–469.
Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.
Sundberg, J., and Gauffin, J. (**1979**). "Waveform and spectrum of the glottal voice source," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Oehman (Academic, London), pp. 301–320.
Švec, J. G., and Pešák, J. (**1994**). "Vocal breaks from the modal to falsetto register," Folia Phoniatr (Basel) **46**, 97–103.
Švec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.
Švec, J. G., 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**, 347–353.
Tembrock, G. (**1996**). *Akustische Kommunikation bei Säugetieren (Acoustic Communication in Mammals)* (Wissenschaftliche Buchgesell, Darmstadt, Germany).
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**2000**). *Principles of Voice Production*, 2nd ed. (National Center for Voice and Speech, Iowa City, IA).
Titze, I. R. (**2006**). *Myoelastic Arodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA).
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.
Titze, I. R., Riede, T., and Popolo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.
Titze, I. R., and Story, B. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.
Tokuda, I., Zemke, M., and Herzel, H. (**2008**). "Biomechanical modeling of voice registers and their transitions," in Proceedings of the 6th International Conference on Voice Physiology and Biomechanics, Tampere, Finland, pp. 98–100.
Tokuda, I. T., Horáček, J., Švec, J. G., and Herzel, H. (**2007**). "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," J. Acoust. Soc. Am. **122**, 519–531.
Vilkman, E., Alku, P., and Laukkanen, A. (**1995**). "Vocal-fold collision mass as a differentiator between registers in the low-pitch range," J. Voice **9**, 66–73.
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," Bioacoustics **9**, 171–196.
Zaccarelli, R., Elemans, C. P. H., Fitch, W. T., and Herzel, H. (**2006**). "Modelling bird songs: Voice onset, overtones and registers," Acta Acust. **92**, 741–748.
Zañartu, M., Mongeau, L., and Wodlicka, G. R. (**2007**). "Influence of acoustic loading on an effective single mass model of the vocal folds," J. Acoust. Soc. Am. **121**, 1119–1129.
Zemke, M. (**2008**). "Biomechanical-aerodynamical modeling of register transitions of the human voice," Diploma thesis, Humboldt University of Berlin, Berlin, Germany.
Zhang, Z. (**2009**). "Characteristics of phonation onset in a two-layer vocal fold model," J. Acoust. Soc. Am. **125**, 1091–1102.
Zhang, Z., Neubauer, J., and Berry, D. A. (**2006**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120**, 1558–1569.

Three-dimensional nature of the glottal jet[a)]

Michael Triep[b)] and Christoph Brücker
Institute of Mechanics and Fluid Dynamics, Chair of Fluid Mechanics and Fluid Machinery, Technische Universität Bergakademie Freiberg, Lampadiusstrasse 4, 09596 Freiberg, Germany

(Received 4 December 2008; revised 17 September 2009; accepted 17 September 2009)

The factors contributing to human voice production are not yet fully understood. Even normal human phonation with a symmetric glottal opening area is still the subject of extensive investigation. Among others, it has already been shown that fluid dynamics has a strong influence on the vocal process. The full characterization of the glottal jet has not been accomplished yet. Time-resolved measurement and visualization of the three-dimensional (3D) flow downstream the human vocal folds are difficult if not impossible to perform *in vivo*. Therefore, it is common to use mechanical and numerical models with a simplified shape and motion profile of the vocal folds. In this article, further results regarding the 3D flow structure obtained in a 3:1 up-scaled dynamic glottis model (cam model) in a water circuit are given, extending earlier work [M. Triep *et al.* (2005). Exp. Fluids **39**, 232–245]. The model mimics the temporal variation in the 3D contour of the glottal gap while water flow reduces the characteristic frequencies by the order of 1/140. The unsteady flow processes downstream of the vocal folds are visualized in slow motion and analyzed in detail via particle imaging techniques. The visualization results show complex 3D flow behavior of lengthwise jet contraction and axis switching. In addition, the time-dependent flow rate during the phonatory oscillation cycle is measured in detail. It is shown that the pressure loss is decreased in the presence of a second constriction downstream of the glottis in form of ventricular folds and it is observed that for this case the jet is stabilized in the divergent phase of the cycle.
 [DOI: 10.1121/1.3299202]

PACS number(s): 43.70.Aj [AH] Pages: 1537–1547

I. INTRODUCTION

The detailed knowledge of flow behavior and real flow processes downstream of the human vocal folds is nowadays subject of extensive study. From their investigations fluid dynamic researchers expect to gain full insight into the flow mechanisms that are important to speech production. In order to do so, the glottal jet has to be described accurately; in particular, three aspects are of immense concern: (a) the three-dimensional (3D) nature of the glottal jet, (b) how variable is the jet behavior from cycle-to-cycle, and (c) what effect do ventricular folds have on glottal flow. Commonly, there are three approaches that are followed for glottal flow investigations: *in vivo* study, the use of *ex vivo*/*in vitro* models, and controlled physical models. The latter has the great advantage that voice production investigations can be carried out in a more controllable environment, such that time-resolving measurements and visualizations of the 3D flow are feasible. These are necessary for the detailed study of the supraglottal development of vortical flow structures, their interaction within the vocal tract, and their impact on flow-induced acoustics.

Important references related to experimental studies involving mechanical physical models of the vocal folds are shown chronologically in Table I. These focused on the study of isolated aspects of the glottal flow.

An important step toward the theory of the glottal physics was first established by Van den Berg (1958) with the myoelastic-aerodynamic theory of voice production. Ishizaka and Matsudaira (1972) developed simplified models for the quantification of the pressure loss across a glottal opening at different subglottal pressures. In order to evidence the self-excited oscillations, silicone vocal folds models have been built; however, these restricted the optical accessibility necessary for detailed flow measurements in the close-up region of the glottis. Furthermore, these models were not appropriate to enforce specific oscillation patterns. Therefore, static models in a transparent test section (Pelorson *et al.*, 1994; Alipour *et al.*, 2002; Zhang *et al.*, 2004; Kucinschi *et al.*, 2006; Erath and Plesniak, 2006a) and models with controllable vocal folds motion (Mongeau *et al.*, 1997; Barney *et al.*, 1999; Zhang *et al.*, 2002b; Triep *et al.*, 2005; Kucinschi *et al.*, 2006; Krane *et al.*, 2007) were employed to study open questions regarding the glottal flow structure. Detailed flow investigations were pursued to study the importance of flow asymmetry on the evolving glottal jet in static two-dimensional (2D) vocal folds models and the evolution of vortical structures downstream the vocal tract (Hirschberg, 1992; Pelorson *et al.*, 1994; Erath and Plesniak, 2006b). Different groups reproduced experimentally an opening and closing motion in lateral direction (Zhang *et al.*, 2002b; Deverge *et al.*, 2003). The results show that the characteristics of the unsteady evolution of the glottal jet differs largely from a quasisteady flow approximation with regard to flow asymmetry and flow mechanisms involving vortical struc-

[a)]This paper is based on a talk presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: michael.triep@imfd.tu-freiberg.de

TABLE I. Experimental fluid-dynamical works on vocal folds mechanical models [only the name of the first author is given: Van den Berg (1958), Ishizaka and Matsudaira (1972), Scherer *et al.* (1983), Gauffin *et al.* (1983), Pelorson *et al.* (1994), Mongeau *et al.* (1997), Barney *et al.* (1999), Alipour *et al.* (2002), Zhang *et al.* (2002a), Agarwal *et al.* (2004), Zhang *et al.* (2004), Neubauer *et al.* (2004), Brücker *et al.* (2004), Triep *et al.* (2005), Krane *et al.* (2004), Krane *et al.* (2007), Alipour and Scherer (2006), Kucinschi *et al.* (2006), Berry *et al.* (2006), Erath and Plesniak (2006b), Drechsel and Thomson (2008), and Bailly *et al.* (2008)] FV=flow visualization, HWA=hot wire anemometry, PT=pressure taps, AM=acoustic measurements, PIV=particle-image velocimetry, and LDV=laser Doppler velocimetry.

	Dimension		Medium		Model				
	2D	3D	Air	Water	Static	Driven	Excited	Main focus	Methods
Van den Berg 1958	×		×		×			Volume flow	Theoretical, PT
Ishizaka 1972	×		×		×		×	Volume flow	Theoretical, PT
Scherer 1983	×		×		×			Volume flow	PT
Gauffin 1983	×		×		×			VFs, volume flow	PT
Pelorson 1994	×		×		×			Separation point, Coanda	FV, HWA, PT
Mongeau 1997		×	×			×		Pulsating jet, unconfined	HWA, PT, FV
Barney 1999	×		×			×		Jet structure, unconfined	HWA, PT, AM
Alipour 2002	×		×		×			Jet turbulence	HWA, PT
Zhang 2002		×	×			×		Pulsating jet in a tube	PT, AM
Agarwal 2004	×		×		×			VFs, 7.5:1 model	PT
Zhang 2004	×		×		×			Confined stationary jets	PT, AM
Neubauer 2004 (2007)		×	×				×	Jet evolution, unconfined	HWA, FV, (DPIV)
Brücker 2004, Triep 2005		×		×		×		Jet evolution, 3:1 model	TR-PIV
Krane 2004 (2007)	×			×		×		Jet evolution, 10:1 model	DPIV
Alipour 2006		×	×				×	Jet turbulence	PT, HWA
Kucinschi 2006	×		×		×	×		Flow rate, scaled-up 7.5:1	PT, FV
Berry 2006	×		×				×	Irregular vibrations	FV, AM, PT
Erath 2006	×		×		×			Pulsating jet, 7.5:1 model	DPIV, HWA, LDV
Drechsel 2008		×	×				×	Jet evolution, VFs	TR-PIV
Bailly 2008	×		×		×		×	VFs, 3:1 model	PT

tures, depending on the motion of the shape and the open quotient of the glottis. At the same time the flow asymmetry is expected to manifest differently in 3D than in the 2D case. An approximate 3D glottis contour during a complete phonatory cycle is known from Hirano *et al.* (1987) who observed a convergent to divergent contour change in the glottis. Up to date there exists no method to detect the complete glottal gap *in vivo* with sufficient spatiotemporal resolution, so that efforts in improving the detection of the vocal folds movement are being undertaken, e.g., by Döllinger and Berry (2006).

The lack of a physical model of the vocal folds with a realistic reproduction of the time-varying 3D motion of the glottis led to the glottal cam model in an earlier study (Triep *et al.*, 2005), which allows the study of the spatiotemporal evolution of the three-dimensional glottal jet flow close to the glottis. This model can produce volume flow waveforms comparable to those produced in real speech. The glottal models of Mongeau *et al.* (1997), Barney *et al.* (1999), Zhang *et al.* (2002b), and Krane *et al.* (2007) demonstrated a similar time-dependent volume flow behavior. The influence of ventricular folds (VFs) on translaryngeal airflow resistance has been studied extensively for steady flow in a static model (Agarwal *et al.*, 2004). It is not clear yet whether this influence holds also for the actual time-varying flow rate profile. Recent works focused on self-oscillating models (Bailly *et al.*, 2008) and canine *ex vivo* models (Alipour *et al.*, 2007). Further studies shall consider the influence of a controlled dynamic motion of the vocal folds.

It is known from technical applications in fluid mechanics that free jets from elliptic orifices experience under certain conditions, a switching of axes. This three-dimensional flow effect has already been observed in a self-oscillating rubber model of the vocal folds (Neubauer *et al.*, 2007; Drechsel and Thomson, 2008) and in excised canine larynx models (Khosla *et al.*, 2007; Khosla *et al.*, 2008). The results obtained in these models suggest at first glance qualitatively similar flow structures as observed in the controllable driven glottal cam model (Triep *et al.*, 2005). It is assumed that the interaction of the glottal jet with the walls downstream the vocal folds will have a strong influence upon the glottal flow behavior; to what extent is still an open issue. Thus, it remains an open question whether the glottal jet through excised canine larynges, which do not include a vocal tract, behaves differently. The present article continues the experimental characterization of the three-dimensional nature of the unsteady glottal jet carried out by Triep *et al.* (2005) using the three-dimensional, scaled-up, driven model of the vocal folds in a water circuit. The model allows an accurate reproduction of the spatiotemporal displacements of the vocal folds during the phonatory cycle. The flow is pressure driven and the resulting time-varying flow rate can be determined. The characteristic time scales of the phonatory cycle in the model are enlarged, such that time-resolved (TR) measurements and visualizations of the glottal flow can be carried out conveniently. The glottal cam model includes a vocal tract with optional insertion of modeled VFs leading to a glottal jet evolution which is focus of discussion. In the pre-

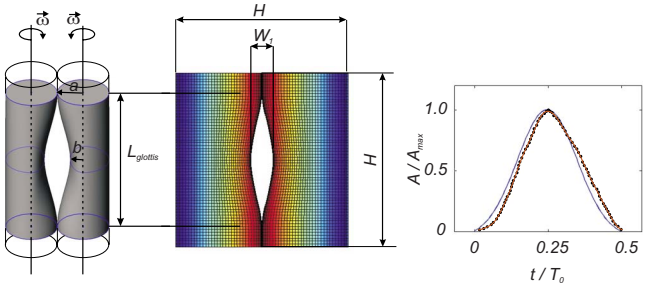


FIG. 1. (Color online) 3D geometry of the cam glottal model in a view from the larynx: (a) semimajor axis, (b) semiminor axis of the ellipse cross sections, and ($\vec{\omega}$) angular velocity (left) and view onto the glottal orifice from downstream in the instant of maximum opening (middle), resulting waveform of the projected normalized glottal area over the cams rotation (right, solid line) and exemplary *in vivo* glottal area measurements taken from Koike *et al.*, 1983, p. 302 (right, dotted line).

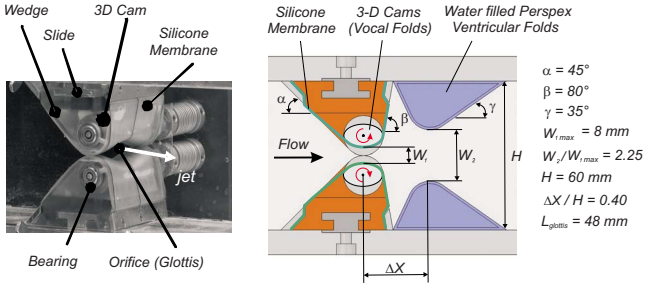


FIG. 2. (Color online) Photograph (left) and midcoronal plane side view of cam model (right, with VFs). The glottis is turned by 90° to the horizontal.

vious work, the sagittal plane jet motion has not been studied in as much detail because the mismatch in the refractive indices between the solid VFs model and the working fluid precluded it. The goal of the present study is to experimentally characterize the gross features of the time-evolving 3D jet flow field. This is done in order to determine whether the observed effect of ventricular folds on the glottal jet, previously studied in static models by Agarwal *et al.* (2003) and more recently also by Bailly *et al.* (2008), would also be observed when the vocal folds move and the 3D jet evolves in time.

The following issues can be addressed with the glottal cam model:

- the 3D characterization of the entire glottal jet including the emerging vortical structures and their interaction in the vocal tract;
- glottal jet evolution, instabilities, and breakdown and its cyclic behavior;
- the character of the flow past the glottis when including VFs in the model;
- detailed flow visualizations of the jet; and
- the shape and integral of the volume velocity waveforms for selected transglottal pressures.

Looking at its capabilities with regard to the flow measurements and the prescription of relevant parameters (pressure head, fundamental frequency, glottal motion profile, glottal orifice shape, and open quotient), the glottal cam model can contribute in the study of the open questions in the glottal jet physics.

Section II briefly describes the driven 3D vocal fold cam model. Thereafter, the experimental setup and the experimental velocimetry and visualization techniques are described. Finally, results of the glottal volume velocity waveform and the unsteady 3D flow field in this model are visualized and discussed for the case without and with model ventricular folds in natural human phonation conditions.

II. THE CAM MODEL AND EXPERIMENTAL SETUP

The cam model has been described in full detail in earlier publications (Brücker *et al.*, 2004; Triep *et al.*, 2005). A short synopsis of the philosophy behind the cam model, the characteristic flow parameters, and the main features of the model are given here. The cam model consists of two counter-rotating 3D shaped driven cams (Fig. 1), each of which is covered with a stretched silicone membrane (Fig. 2). The global parameters of the contour of the model glottis and its motion (see Fig. 3), which are important for the flow evolution, show the most realistic basic features in comparison to other medical observations. Especially the convergent-divergent shape change in the glottal opening during the cycle is replicated similar to that observed by Hirano *et al.* (1987). In addition, the elliptic cross section (see Fig. 1 and glottograms from literature) is very similar. Finally, the waveform is very close to what can be found in literature: (1) for volume velocity waveform, see Sulter and Wit, 1996, and (2) for glottal area waveform, see Koike *et al.*, 1983. Note that the membrane motion generated by the cam model comprises a vertical (streamwise) and transversal (crosswise) component.

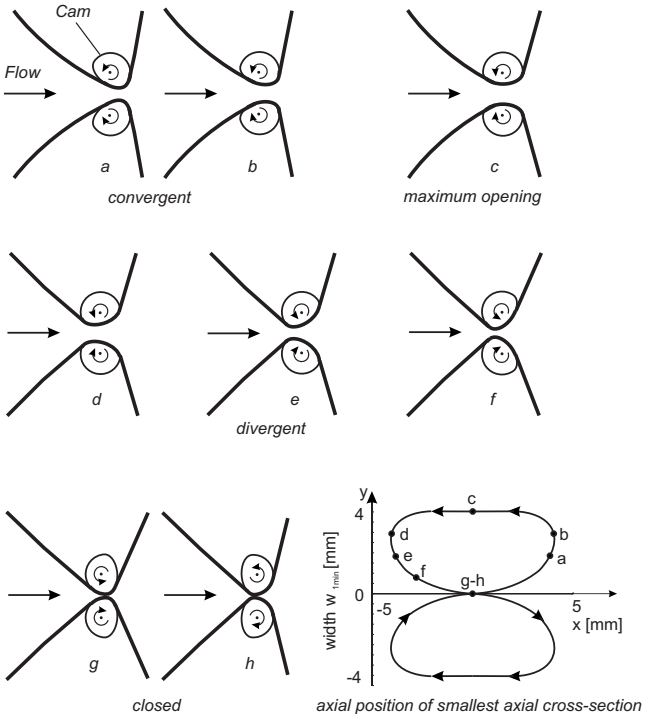


FIG. 3. Variation in the axial cross section of the glottis profile in the midcoronal plane during the glottal cycle and curve of the minimum gap locus over the cycle (bottom, right) [from (a) t/T_0=0.15 until (f) t/T_0 =0.40 in steps of $\Delta t/T_0$=0.05, (g) t/T_0=0.50, and (h) t/T_0=1.0].

Changing the 3D contours (Fig. 1), the rotational speed, or the material of the cams allows to investigate several different phonatory situations. In this paper, similar cams as used in the earlier study which were designed for the reproduction of the natural phonatory oscillation are employed. Besides geometric similarity, the Reynolds (Re) and Strouhal (Sr) numbers are similar for the real situation and in the model experiment. The order of magnitude of the most important characteristic numbers considered here is as follows:

- $\mathrm{Re}_W = u_{\mathrm{mean}} W_{1\,\mathrm{max}} / \nu \approx O(10^3)$ and
- $\mathrm{Sr} = f_0 W_{1\,\mathrm{max}} / u_{\mathrm{mean}} \approx O(10^{-2})$,

where u_{mean} is the typical mean flow velocity in the original glottis during phonation (Pelorson *et al.*, 1994), c is the velocity of sound in air, f_0 is the fundamental frequency of the glottal cycle, $W_{1\,\mathrm{max}}$ is the maximum glottal opening, ν is the kinematic viscosity, and ρ is the density of the fluid. The fundamental frequency $f_0 = 1/T_0$ of the model is in the order of 1 Hz which corresponds to 135 Hz in human voice. Furthermore, the ratio between the dynamic pressures of the model and the human voice leads to a pressure head in the model that is half the value of that prevailing in human voice. The Mach number which is estimated as $\mathrm{Ma} = u_{\mathrm{mean}}/c \approx O(10^{-1})$ is small so that the flow during natural phonation can be considered to be incompressible.

The vocal folds module is placed in the test section of a U-shaped water flow channel arrangement with a column at both ends for imposing the pressure head (Fig. 4). The downstream column includes an overflow with a sharp edge, whereas the inlet column comprises a Styrofoam disk. In general, a free surface of the water column is prone to oscillations resulting from pressure waves at the opening and closing of the glottis. In order to ensure an almost oscillation-free surface, a lightweight Styrofoam disk is placed on the water column. This extension to the previous setup allows a more precise evaluation of the flow velocity waveform measurements. This method of volume flow measurement adds an important advantageous feature of this unique cam model and experimental setup. All components of the channel are made of Perspex, so that optical access is guaranteed for whole plane flow measurements, in particular, in the region close to the glottis and in the Morgagni's ventricle between vocal and VFs. Geometric data of the VFs are according to laminagraphic tracing measurements from Agarwal *et al.*, 2003. Since the refractive index of the Perspex and the working fluid are not matched, the VFs were manufactured of thin material and filled with the working fluid in order to minimize the image distortions.

The glottal cycle in this study has an open quotient of 50% which means that the membranes remain in the closed state for half the period. This is achieved by contouring the cams with a semicircular arc representing the closed phase of 50% of the cycle (Fig. 3). The use of water as the carrier fluid and the upscaling of the glottal model allows us to visualize the flow in slow motion.

The volume flow waveform can be measured accurately by detecting the temporal vertical motion of the disk on the inlet water column reservoir by a microscopic lens/high-speed camera setup and image processing. The flow rate profile is measured for different driving pressure heads in order to study the pressure loss in the near-glottal region. Furthermore, the volume flow waveform is measured for a configuration without and with the VFs for the same initial pressure head condition and driving speed of the cams.

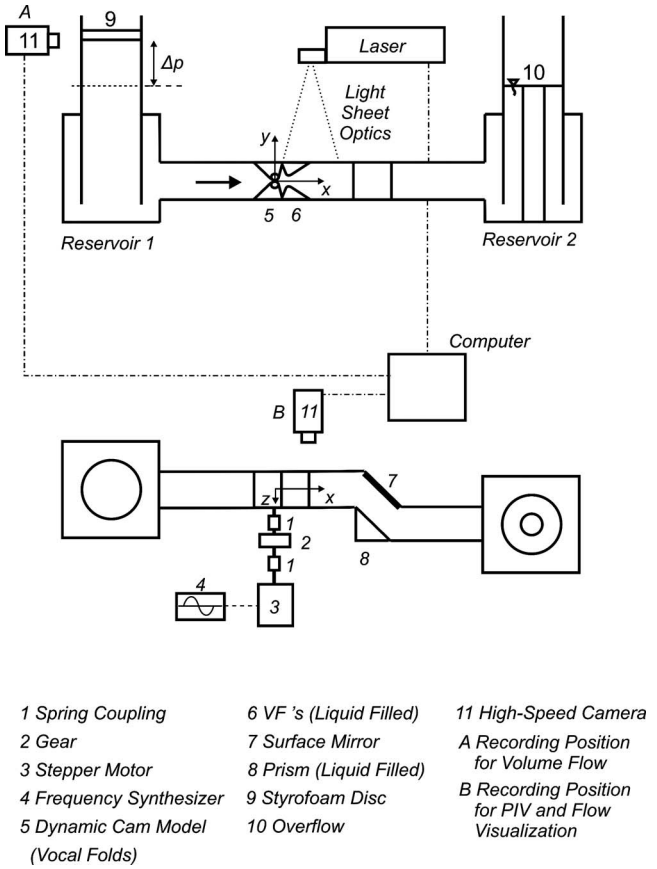


FIG. 4. Front view (top) and top view (bottom) of the experimental setup. The origin of the coordinate system is placed at the center of the glottis; the x-y plane is named coronal plane, the x-z plane is named sagittal plane, and the y-z plane is named transverse plane.

TR flow measurements are carried out downstream of the model glottis by means of high-speed particle-image velocimetry (PIV). The TR-PIV measurements utilize a FASTCAM APX-RS camera (Photron Inc., San Diego, CA, 3000 fps, 1024×1024 pixels) in combination with a Pegasus laser system (New Wave Inc., Fremont, CA). The fluid is seeded with particles with a mean diameter of 20 μm.

In order to provide further important information about the glottal flow in the study of the main issues listed in the Introduction, time-resolved flow visualizations are made. Due to the full optical accessibility into the near-glottal region, the visualization of the glottal flow is accomplished by introducing fluorescent Rhodamine dye homogeneously into the channel downstream of the glottal cam model. In this way, dye-free fluid that passes the glottis becomes visible in the emerging jet (contrast method) which is then recorded in the light sheet with the FASTCAM APX-RS camera. Hereby, we obtain a beneficial view of the overall flow and even of the jet edge instability structures. For all aforementioned measurements, the upstream water column is filled to a level slightly higher than that needed to impose the pressure head

TABLE II. Summary of the experiments.

Type of experiment	Glottal orifice shape	Inclusion of ventricular folds	Open quotient OQ	Physiological pressure head Δp (cm H_2O)	Figure no.
Volume flow	Elliptical	No/yes	0.5	6, 10, 14	5
PIV	Elliptical	No	0.5	6	6
PIV	Elliptical	Yes	0.5	6	7
Visualization	Elliptical	No	0.5	6	11
Visualization	Elliptical	Yes	0.5	6	12
Visualization	Rectangular	Yes	0.5	6	13
PIV	Elliptical	No	0.5	6, 14	14 and 15

across the glottis. For these measurements always the second glottal cycle is observed, and the relative difference in the pressure head for successive measurements is less than 1%. For the purpose of cyclic PIV measurements, an overflow is also introduced in the upstream water column, so that the driving pressure is kept constant for each cycle. The measurements which will be presented in Sec. III are detailed in Table II. The PIV measurements are necessary for the extraction of quantitative data from the flow field while the dye visualizations are rather of qualitative character. Using the method of PIV, the spatial resolution is restricted by the velocity analysis procedure which averages over a small interrogation area that should include at least seven pairs of seeding particles. By contrast the dye visualizations are able to give a very sharp view of the glottal jet edge.

III. RESULTS

A. Volume velocity waveform

The measured results of the volume flow waveform through the glottis model are shown in Fig. 5 for three different pressure heads across the glottis (Δp=6 cm H_2O, Δp=10 cm H_2O, and Δp=14 cm H_2O). All quantities are given as values for real natural phonation. The volume flow waveform is normalized with the friction-free discharge velocity from Torricelli's theorem and the cross-sectional area of the model vocal tract $A_{channel}$.

The shape of the measured flow velocity waveforms look very similar to those found in literature (Cranen and Boves, 1988; White, 1992; Sulter and Wit, 1996). The comparison of the volume flow waveforms for the case without and with model VFs clearly show that the pressure loss is reduced in the presence of VFs, since all curves for this case are always higher than for the cases without. The difference between the integral values is in the order of 10%. The curves show the typical skewing to right, so that the instant within the glottal cycle for maximal volume rate is delayed with respect to the instant of maximal glottal opening (Fig. 1). Note that the glottal area function $A_{glottis}(t)$ in this study is symmetric and the open quotient is 50%, so that the glottis is fully open at 25% of the cycle after glottis opening.

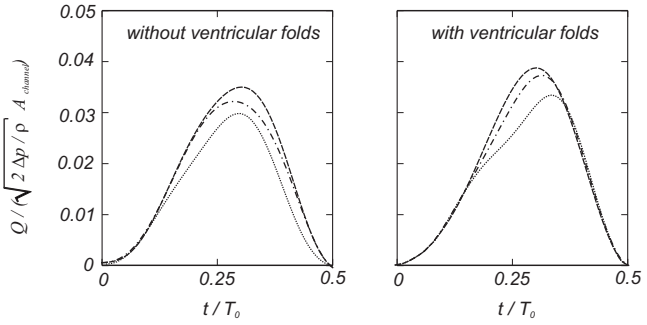


FIG. 5. Measured volume flow rates Q for imposed transglottal pressure heads of Δp=6 cm H_2O (dotted line), Δp=10 cm H_2O (chain dotted line), and Δp=14 cm H_2O (dashed line). The volume flow $Q(t)$ is made dimensionless with the loss-free steady flow rate at the given pressure head.

For the case with VFs, the instant of maximum glottal flow is delayed to later cycle instants for decreasing pressure heads. This tendency is not clearly observed for the case without VFs, where the maximum glottal flow occurs basically at the same instant in the cycle. The effect of strongly abducted VFs on glottal flow, observed by Scherer *et al.* (1983) and Agarwal *et al.* (2004) for steady flows through static geometries, is also observed in the present case, in which the model vocal fold walls move and the flow is time varying. The changes observed in the glottal flow jet, due to the presence of the VFs, will be discussed in more detail in Sec. III B. These changes support the notion that the VFs modify glottal jet entrainment such that the volume flow per unit transglottal pressure is enhanced.

B. Flow evolution

The generation and breakdown of the unsteady jet through the moving glottis is well described in space and time over 1 cycle and characteristic details of the flow are extracted and studied. For a visual representation of the different cross sections of the high-speed PIV measurements and flow visualization, see the coordinate system depicted in Fig. 4. Measurements are carried out in the midcoronal (x-y) plane and in the midsagittal (x-z) plane without and with VFs. Further results from measurements in one exemplary transversal (y-z) plane are also shown.

Figure 6 shows the velocity vector plots in the midcoronal and midsagittal planes of the glottal model without VFs at six characteristic instants of the open phase of the glottal cycle, taken from a single high-speed sequence: t_1/T_0=0.15 (a) in Δt steps of 0.05 up to t_6/T_0=0.40 (f). In the following description the letters (a)–(f) are used when referring to these six instants. The glottal jet front plume and jet edge can be deduced from the velocity vector lengths while considering their actual point of action. In order to better accentuate the

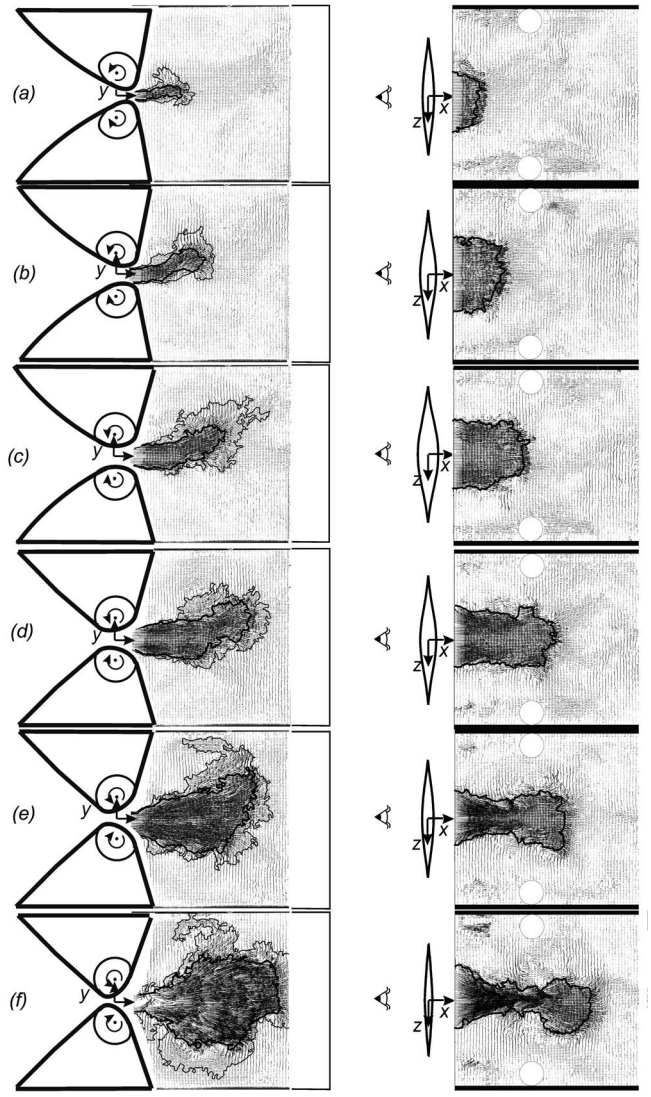


FIG. 6. Flow field at instants (a) $t/T_0=0.15$ to (f) $t/T_0=0.40$ in steps of $\Delta t/T_0=0.05$ after glottis opening (without VFs) in the midcoronal plane (left) and midsagittal plane (right) of the channel for $\Delta p=6$ cm H_2O: velocity vector plots with isocontours of in-plane speed as values in real natural phonation (3.5 m/s—thin solid line and 10 m/s—thick solid line).

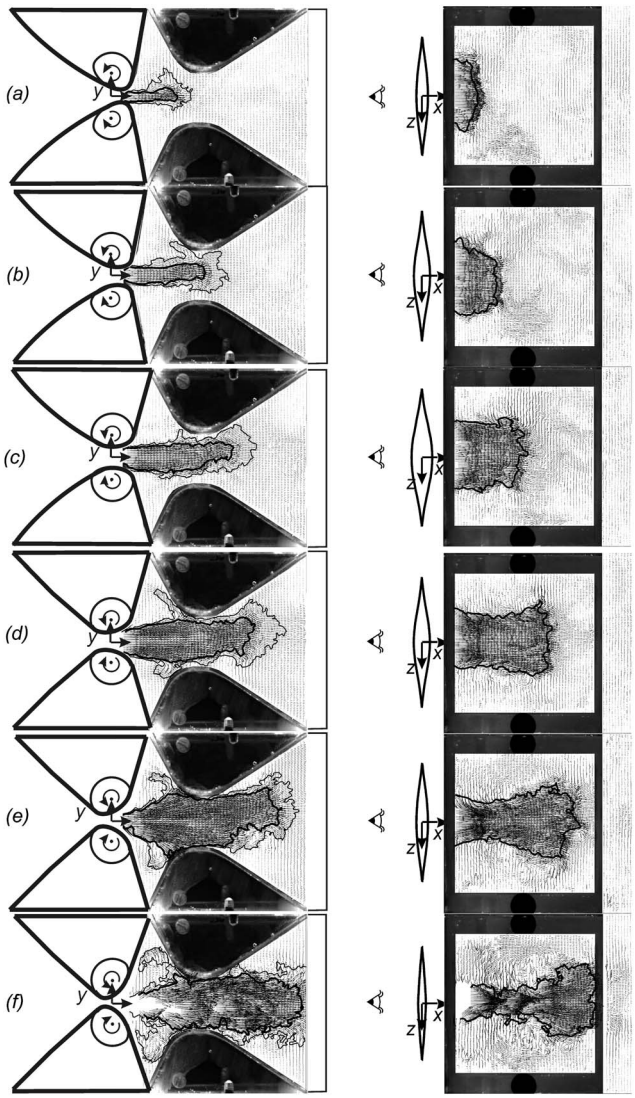


FIG. 7. Flow field with VFs in the midcoronal plane (left) and midsagittal plane (right) of the channel at same instants as in Fig. 6 for $\Delta p=6$ cm H_2O.

jet, the isocontours of the in-plane speeds of 3.5 and 10 m/s are also shown in the midcoronal planes. For the sake of clarity in the midsagittal planes, only the isocontour of 10 m/s is displayed.

After opening of the glottis, the jet exits the glottal gap in this cycle almost aligned with the vocal tract downstream axis. Small deviations are observed in the front plume of the jet. Measurements of several successive glottal cycles show that there are cycles where the jet emerges symmetrically and that this behavior occurs randomly. In this initial stage of the jet evolution, the jet approximately retains the geometrical shape and orientation of the glottal orifice [Figs. 6(a)–6(c)]. The shape of the jet cross section [Fig. 6(c)] remains nearly elliptical as in Fig. 6(a). In the follow-up, both midplane plots clearly reveal the presence of a lengthwise vena contracta. A careful inspection of the jet width in the midsagittal plane shows a somewhat smaller width than at the initial state after glottis opening. This trend continues for the whole phase in which the glottis is open, meaning that the vena contracta effect in this plane becomes more pronounced during the further evolution of the jet [Figs. 6(d)–6(f)]. Maximum volume flow is reached between instants (d) and (e). In this case without VFs, the jet is allowed to spread in crosswise direction in the vocal tract during its emerging and evolving stages: the jet front almost occupies the complete channel height in the midcoronal plane.

Figure 7 shows the flow results in the same midsagittal and midcoronal planes for the case with VFs. This helps to describe the influence of the VFs upon the glottal jet. After opening of the glottis, the jet emerges and is immediately attached in this cycle to the upper wall of the VFs [Figs. 7(a) and 7(b)]. This asymmetry in the flow has been already observed in the previous study by Brücker *et al.* (2004) and later by other groups (Kucinschi *et al.*, 2006; Krane *et al.*, 2007). At maximum glottal opening [Fig. 7(c)], the jet is still attached to the upper wall of the VFs. The lengthwise vena contracta is also clearly observed in both midplane plots for the case with VFs during the closing phase, with a similar

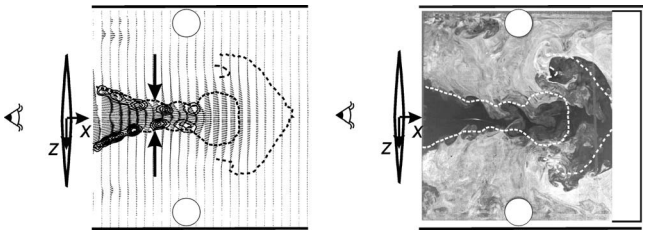


FIG. 8. Flow field at $t/T_0=0.40$ after glottis opening (without VFs) in the midsagittal plane of the channel for $\Delta p=6$ cm H_2O. Left: velocity vector plot with isocontours of in-plane speed as values in real natural phonation of 3.5 and 10 m/s (dashed lines) and isocontours of lengthwise vorticity from 0.02 to 0.2 (solid lines). Right: flow visualization image including isocontours of in-plane speed from left image.

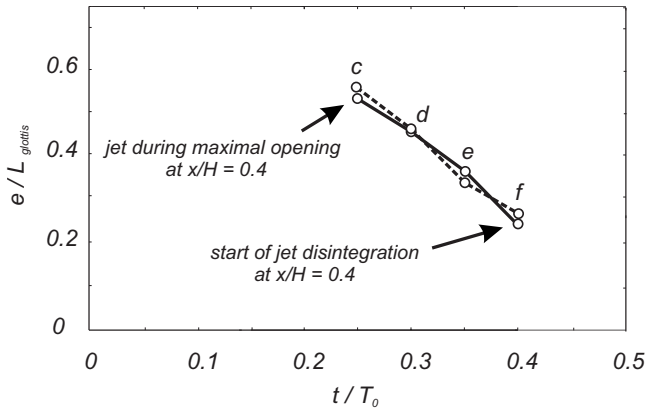


FIG. 9. Temporal change in jet vena contracta width for the case without VFs in the midsagittal plane at $x/H=0.4$ for $\Delta p=6$ cm H_2O from PIV data (dashed line) and flow visualizations (solid line).

temporal evolution in the midsagittal plane as in the case without VFs. Meanwhile, one can see that the jet starts to detach from the downstream facing walls of the VFs [Figs. 7(d) and 7(e)] and enters the vocal tract in axial direction [Fig. 7(f)]. This suggests one of the functions of the second constriction in the channel between the VFs to focus the jet in axial direction. Furthermore, recirculations are formed in the Morgagni's ventricles during the glottal cycle. Analysis of several successive cycles show that the characteristic dynamics of the jet repeat similarly to the previous cycle. However, the side to which the jet first attaches to the VFs after glottis opening seems to occur randomly. The measurements suggest that this is linked to the influence of flow structures left in the vocal tract by the jet breakdown in the previous cycle. Therefore, it is expected to see a stronger influence of the shut-off phase of the jet onto the next cycle for increased open quotients.

For both cases, without and with VFs, instabilities occur already in the early stages of the jet formation in form of staggered shear-layer vortices. The jet flow with its small to moderate velocities in this convergent opening phase of the glottal cycle is accelerated further in the divergent closing phase. In the late stage, a meandering of the jet tail with high velocities is observed, while the jet front—a plumelike structure—travels further unaffected. This finally results in a smoothly shrinking and rupture of the tail at glottis closure, which leaves the jet front plume as an independent bulk of momentum. Further disintegration of the jet plume appears to be of strong dissipative nature. The experiments with VFs show that the VFs prevent a growth of the jet edge instabilities. The jet core is smoothly directed toward the sidewalls of the ventricular walls, where the jet builds up a new boundary layer. It is clear from the experiments without and with the VFs that the second constriction built by the VFs additionally stabilizes the jet.

The obvious difference of the jet cross-section in the different transverse planes demonstrates the 3D nature of the jet flow which is in addition changing over time. First, the flow in the midsagittal plane of the glottal model without VFs is considered. A coefficient of contraction is defined here as the ratio between the contraction width e at the downstream position $x/H=0.4$ and the instantaneous glottal length, L_{glottis}, so that the phenomenon of the lengthwise vena contracta is quantified. For the determination of the contraction width e, consider Fig. 8 (left). This figure shows exemplarily for the case without VFs the velocity field in the midsagittal plane obtained by PIV at instant (f). Note that for the sake of clarity, only each fifth velocity vector in x-direction is shown. In addition, the isocontours of the in-plane speed of 10 m/s (thick dashed line) and partially of 3.5 m/s (thin dashed line) are illustrated. The closed solid isocontour lines designate regions of concentrated crosswise vorticity in levels with values ranging between 0.02 and 0.2 s^{-1}. A convenient way to define the jet edge is to connect the points of maximum vorticity in the shear-layer vortices. Figure 8 (left) demonstrates that the isocontour of the in-plane speed of 10 m/s is appropriate to define the jet edge and by this means to determine the contraction width e of the glottal jet (indicated here by the arrows at position $x/H=0.4$). The coefficient of contraction is plotted in Fig. 9 for the instants (c)–(f) in the glottal cycle. An approximate linearly increasing contraction (decreasing e/L_{glottis}) is observed for the late stage of the glottis motion. For the sake of completeness, the isocontours of speed have been transferred to the corresponding flow visualization image in Fig. 8 (right) affirming that the jet in the PIV measurements and in the flow visualizations with dye shows a similar behavior. As already mentioned, the vena contracta phenomenon is enhanced over the cycle, finally leading to a thin meandering high-speed long-tailed structure in the closing phase. This is diametrically opposite to the observed jet evolution in the midcoronal plane. While at the beginning the jet width in this plane is rather small, it continuously increases in the closing phase, where the downstream glottis contour changes from a parallel to a divergent shape [Figs. 3(c) and 3(d)]. To further investigate this behavior, a measuring plane in the cross-sectional transverse plane is arranged at $(x/H=0.4)$. The PIV measurements were analyzed in the way to discriminate the jet core in the radial cross section from the ambient fluid. Again, isocontours of speed delineate the border of the jet core. These contours are shown in Fig. 10 over successive instants in the cycle. While the cross section has an elliptic contour in the same orientation as the glottal orifice until instant (d), it switches its aspect ratio at the end of the open cycle [Figs. 3(e) and 3(f)]. Note that this behavior results from the imposed motion of the orifice over time and therefore, it can be titled as in-cycle temporal axis switching of

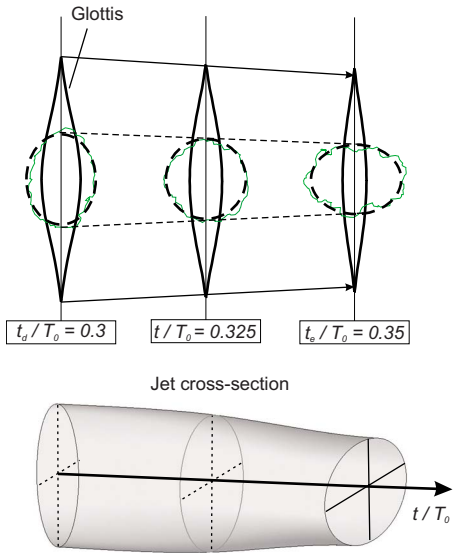


FIG. 10. (Color online) Approximate shape of temporal evolution of the average jet cross section (contour of constant in-plane velocity) in the plane $x/H=0.4$.

the jet. The phenomenon of axis switching in jets through static elliptic orifices has been studied in the past by Hussain and Hussain (1989).

The flow visualizations shown in Fig. 11 (case without VFs) and Fig. 12 (case with VFs) confirm the PIV results. The left column shows the midcoronal plane, and the right column shows the midsagittal plane of the vocal tract. In Fig. 11 the six instants (a)–(f) are presented from top to bottom. In Fig. 12, only the instants (c)–(e) are shown. After jet plume formation, characteristic coherent structures become visible in the jet. It has been observed that these jet structures and their interactions, including vortex pairing, are reproducible from cycle to cycle. Jet axis switching is also evident from the visualizations. The strong lengthwise vena contracta, along with the meandering and decay of the jet, is characteristic in the second half of the open part of the cycle. The small scale structures of the jet edge shear-layer instability are of the order of 1/40 of the channel height and are focus of current studies with measuring techniques with high resolution such as PIV. The coefficients of contraction obtained from these flow visualizations for the instants (c)–(f) in the glottal cycle have been added in the plot in Fig. 9 (solid line). The values show good agreement with the data from the PIV measurements. The existent deviations are first due to different glottal cycles used in the measurements and second due to differences in the spatial resolution of PIV and the flow visualizations. The limited spatial resolution in PIV leads to a frazzle of the isocontours of the in-plane speed. By contrast the dye visualizations give a detailed view of the jet edge instabilities and Kelvin–Helmholtz vortices.

In this section, the flow downstream the dynamic elliptical glottal orifice without and with VFs has been shown to be highly unsteady and of three-dimensional nature. The results show the following: jet axis switching, a strong lengthwise vena contracta which evolves in time, jet edge instabilities, and jet breakdown. Repeated measurements have shown

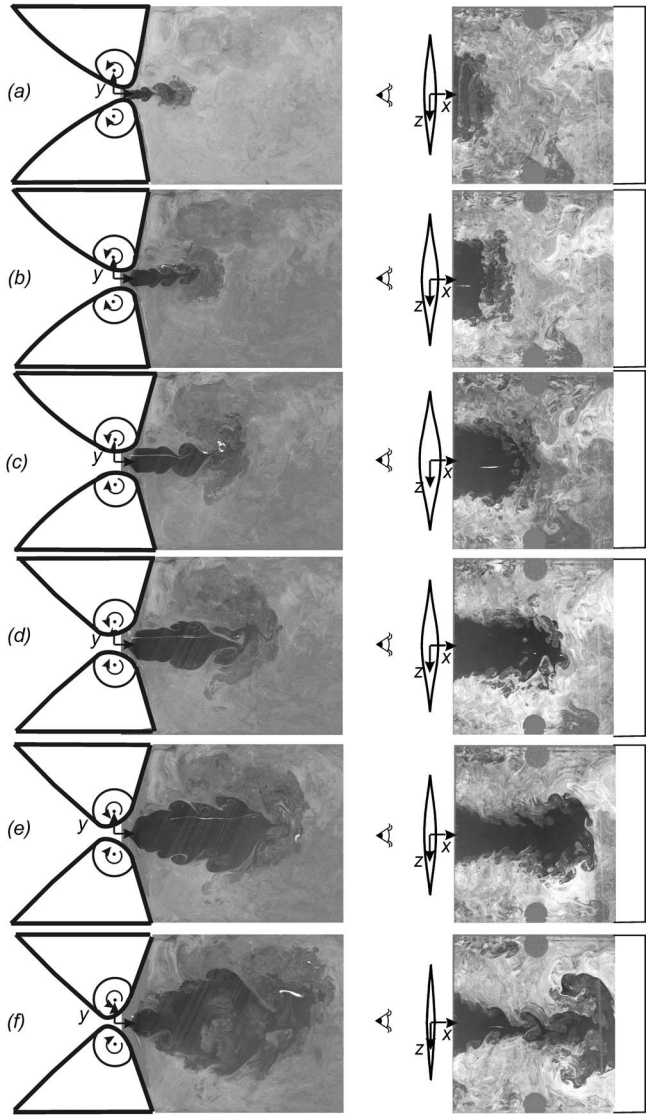


FIG. 11. Visualization of the glottal flow (without VFs) in the midcoronal plane (left) and midsagittal plane (right) of the channel for $\Delta p = 6$ cm H_2O for instants as in Fig. 6.

that the jet behavior is consistent from cycle to cycle, despite some cycle-to-cycle variations mainly at the early stages of the jet formation.

C. Influence of glottal orifice shape and driving pressure head

The character and behavior of the 3D jet past the model glottis depend on the glottal orifice shape. Mainly in the sagittal plane, the flow structures downstream of the cam model are expected to develop according to the 3D contour of the glottal orifice shape. In order to investigate this influence, additional measurements were performed for the jet in the midsagittal plane downstream of a glottal orifice with a rectangular cross section but a similar dynamic behavior of the opening and closing contours, compare Fig. 3. The normalized glottal area time function for the rectangular opening has been chosen to be similar to that of the elliptical opening shown in Fig. 1. The flow in the midsagittal plane for the instants (c)–(e) is shown in Fig. 13 in a view through

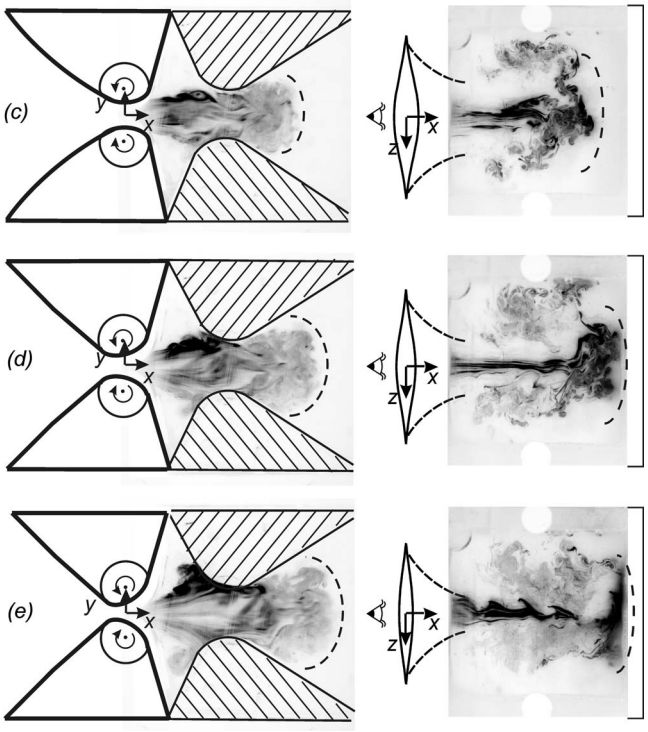


FIG. 12. Visualization of the glottal flow (with VFs) in the midcoronal plane (left) and midsagittal plane (right) of the channel for $\Delta p=6$ cm H_2O for instants as in Fig. 6. The dashed lines indicate the location of the jet front.

the liquid-filled modeled VFs. For the rectangular glottal opening, the contraction of the jet in this plane is constant throughout the open phase of the glottal cycle and is smaller than in the case of the elliptical glottal orifice in Fig. 12. These visualizations evidence that the strong jet contraction in the elliptical glottal orifice case is the effect of the shrinking of the glottis length L_{glottis} during the closing phase.

Next, the influence of the driving pressure head is shown in Fig. 14. The flow in the midsagittal plane (without VFs) for the instant of maximum opening is compared for the pressure heads of $\Delta p=6$ cm H_2O and $\Delta p=14$ cm H_2O. These results are velocity fields phase-averaged over 100 cycles at the same cycle instant. The cyclic variations are exemplarily shown in the instantaneous velocity fields for three different cycles in Fig. 15. These evidence the random tendency (slightly upward, straight or slightly downward) of the plume front of the emerging glottal jet from cycle to cycle. For $\Delta p=14$ cm H_2O, the jet front has traveled almost the double distance further downstream into the vocal tract than for $\Delta p=6$ cm H_2O. In general, the jet shows a similar behavior from cycle to cycle. In the starting acceleration stage of the flow, the low speed jet front is sensitive to any disturbances in the remaining flow in the vocal tract and to smallest inaccuracies in the cams that form the vocal folds. These effects together determine the starting behavior of the jet front. For increased transglottal pressures and in the later stages in its evolution, the jet is stabilized due to the higher velocities.

IV. SUMMARY AND OUTLOOK

The complex unsteady 3D vortex dynamics in the vocal tract was studied in a 3:1 upscaled dynamical model of the human vocal folds (Fig. 2). Volume flow velocity waveform measurements and flow visualizations for normal phonation conditions have been carried out. The modular construction of the setup allowed analyzing the flow in the absence and presence of modeled VFs, which build a second constriction in the test channel. A realistic volume velocity waveform as observed in real speech is obtained for the prescribed transglottal pressure, fundamental frequency, open quotient, and realistic glottal motion. Characteristic details of the evolution and redistribution of vortices could be detected qualitatively and quantitatively.

The experiments reveal a clear picture of the 3D and

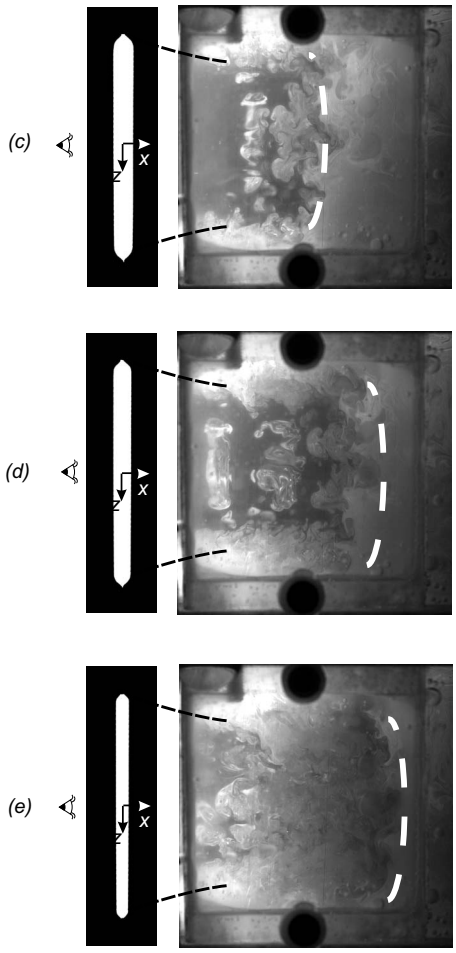


FIG. 13. Visualization of the glottal flow (with VFs) in the midsagittal plane of the channel for $\Delta p=6$ cm H_2O and a rectangular opening shape of the glottis for the instants (c)–(e). The black dashed lines indicate the phase-invariant constant contraction of the jet, while the white dashed lines indicate the location of the jet front.

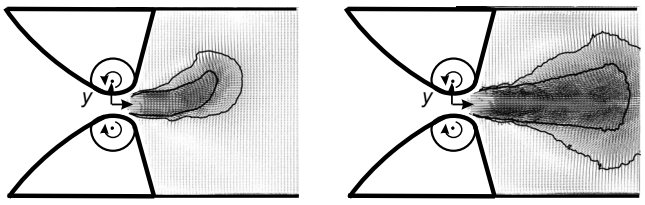


FIG. 14. Mean flow field at instant (c) $t/T_0=0.25$ after glottis opening (without VFs) in the midcoronal plane of the channel for $\Delta p=6$ cm H_2O (left) and $\Delta p=14$ cm H_2O (right): velocity vector plots with isocontours of in-plane speed (3.5 and 10 m/s).

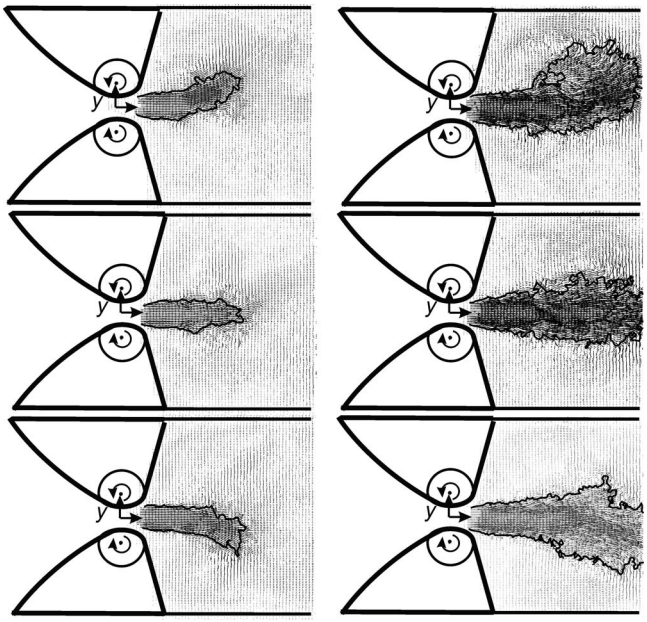


FIG. 15. Instantaneous flow field at instant (c) $t/T_0=0.25$ after glottis opening exemplarily in three different glottal cycles: (without VFs) in the midcoronal plane of the channel for $\Delta p=6$ cm H_2O (left) and $\Delta p=14$ cm H_2O (right): velocity vector plots with isocontours of in-plane speed 10 m/s.

unsteady nature of the glottal jet in the glottal cam model.

(a) The flow rate measurements in the cam model indicate that the maximum volume flux occurs later than the maximum opening of the glottis. Therefore, convective flow acceleration in the glottis is still present in the closing phase and maximum flow velocities occur later in the cycle while the glottis is in the closing phase. In addition, the measurements demonstrate a reduction in the dynamic transglottal pressure loss in the case of a vocal tract with VFs.

(b) There is a lengthwise vena contracta, which increases with time up to the end of the cycle. Thus the developing jet does not have its maximum width in the lengthwise direction but rather the opposite. This is the consequence of two contributing factors: first, flow through an elliptic orifice in general leads to the so called axis switching of the jet, which is observed in these dynamic experiments, too. Second, with the beginning of glottis closure, the lengthwise scale of the glottal orifice shrinks in the same manner as the width of the glottal gap; therefore, flow is forced to redirect toward the center of the glottal axis. Simultaneously, the cross section of the glottis has changed to the divergent form; thus, the flow can still expand in crosswise direction. These dynamic features of the jet strongly depend on the glottal motion pattern and glottal orifice shape, as the comparison with a rectangular orifice in dynamic mode could show.

(c) No considerable asymmetric lateral or transversal deflection of the glottal jet and no final attachment to one of the side walls of the vocal tract is observed. Instead, the jet exits the glottal gap mainly aligned with the channel axis. Some small deviations could be observed in the regions of the front plume of the jet. This observation holds for the complete buildup and decay of the jet. With increasing pressure head, the jet front is redirected more with the centerline direction of the channel axis. The fluctuations at lower pressure heads seem to result from small inaccuracies of the cams, which may not be perfectly symmetric and which affect the flow only for smaller velocities. The results show the important differences in the flow of a dynamic glottal motion in comparison with the flow results in a static orifice. With the given open quotient of 0.5 and the pressure heads investigated, it seems that the characteristic time scale of buildup and decay of the jet is not long enough to establish any energetic large scale vortices downstream in the vocal tract, which might support the jet deflection. Again, the presence of the VFs further reduces any asymmetric features of the early stage of the jet front and redirects the jet into the axial direction.

(d) The dynamic jet maintains shear-layer instabilities and vortex rollup during the complete cycle of buildup and decay. These vortices interact with the walls on the VFs and may contribute to vortex-induced sound in the acoustic spectrum of the primary sound source generated in the glottis.

(e) Due to the delay of maximum volume flux with respect to maximum glottal opening, one would expect a rather sharp jet breakdown. The measurements on the other hand show a smooth disintegration of the jet at the late stage of glottal closure due to the three-dimensional closure with a smooth closing of the contact line similar to a closing two-slider zipper from both sides of the lateral edges of the glottis.

These results hold for the given conditions in the glottal cam model, prescribed in the form of the pressure head, the fundamental frequency, the glottal motion profile, the glottal orifice shape, and the open quotient. The sensitivity of the flow field to changed parameters, especially to the way the glottal profile and waveform is replicated in the experiments, will be studied in more detail in the near future. Other glottal opening types especially nonsymmetric openings may lead to strongly different flow fields.

The experimental results are used for input and comparison with numerical simulation of our group with exactly the same initial and boundary conditions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of this project by the Deutsche Forschungsgemeinschaft in the Research Group DFG FOR 894 under Contract No. Br1494/13-1. In addition, we wish to thank Mr. Clemens Kirmse for the support in the experimental measurements and the reviewers for their critical but helpful comments.

Agarwal, M., Scherer, R., and Hollien, H. (**2003**). "The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings," J. Voice **17**, 97–113.

Agarwal, M., Scherer, R., and Witt, K. D. (**2004**). "The effects of the false vocal folds on translaryngeal airflow resistance," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille, France.

Alipour, F., Jaiswal, S., and Finnegan, E. (**2007**). "Aerodynamic and acoustic effects of false vocal folds and epiglottis in excised larynx models," Ann. Otol. Rhinol. Laryngol. **116**, 135–44.
Alipour, F., Montequin, D., and Scherer, R. (**2002**). "Asymmetric vocal fold vibrations," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Denver, CO.
Alipour, F., and Scherer, R. (**2006**). "Characterizing glottal jet turbulence," J. Acoust. Soc. Am. **119**, 1063–1073.
Bailly, L., Pelorson, X., Heinrich, N., and Ruty, N. (**2008**). "Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation," J. Acoust. Soc. Am. **124**, 3296–3308.
Barney, A., Shadle, C., and Davies, P. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract," J. Acoust. Soc. Am. **105**, 444–455.
Berry, D., Zhang, Z., and Neubauer, J. (**2006**). "Mechanisms of irregular vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**, EL36–EL42.
Brücker, C., Triep, M., and Kob, M. (**2004**). "Study of the vortex dynamics in a mechanical model of the vocal folds using particle-image velocimetry," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille, France, pp. 11–17.
Cranen, B., and Boves, L. (**1988**). "On the measurement of glottal flow," J. Acoust. Soc. Am. **84**, 888–900.
Deverge, M., Pelorson, X., Vilain, C., Lagre, P., Chentouf, F., Willlems, J., and Hirschberg, A. (**2003**). "Influence of collision on the flow through in-vitro rigid models of the vocal folds," J. Acoust. Soc. Am. **114**, 3354–3362.
Döllinger, M., and Berry, D. (**2006**). "Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation," J. Voice **20**, 401–413.
Drechsel, J., and Thomson, S. (**2008**). "Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model," J. Acoust. Soc. Am. **123**, 4434–4445.
Erath, B., and Plesniak, M. (**2006a**). "An investigation of bimodal jet trajectory in flow through scaled models of the human vocal tract," Exp. Fluids **40**, 683–696.
Erath, B., and Plesniak, M. (**2006b**). "The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds," J. Acoust. Soc. Am. **120**, 1000–1011.
Gauffin, J., Binh, N., Ananthapadmanabha, T., and Fant, G. (**1983**). "Glottal geometry and volume waveform," *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (College Hill Press, San Diego, CA), pp. 194–201.
Hirano, M., Yoshida, T., and Kurita, S. (**1987**). "Anatomy and behavior of the vocal process," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and K. Harris (College Hill Press, Boston, MA), pp. 1–13.
Hirschberg, A. (**1992**). "Some fluid dynamic aspects of speech," Bulletin de la Communication Parlée **2**, 7–30.
Hussain, F., and Hussain, H. (**1989**). "Elliptic jets. Part 1. Characteristics of unexcited and excited jets," J. Fluid Mech. **208**, 257–320.
Ishizaka, K., and Matsudaira, M. (**1972**). "Fluid mechanical considerations of vocal fold vibration," SCRL (Speech Communication Research Laboratory) Monograph No. 8, Santa Barbara, CA.
Khosla, S., Muruguppan, S., Gutmark, E., and Scherer, R. (**2007**). "Vortical flow field during phonation in an excised canine larynx model," Ann. Otol. Rhinol. Laryngol. **116**, 217–238.
Khosla, S., Muruguppan, S., Lakhamraju, R., and Gutmark, E. (**2008**). "Using particle image velocimetry to measure anterior-posterior velocity gradients in the excised canine larynx model," Ann. Otol. Rhinol. Laryngol. **117**, 134–144.
Koike, Y., Imaizumi, S., Kitano, Y., Kawasaki, H., and Hirano, M. (**1983**). "Glottal area time function and supraglottal pressure variation," *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (College Hill Press, San Diego, CA), pp. 300–306.
Krane, M., Barry, M., and Wei, T. (**2004**). "Flow measurements in a scaled-up vocal folds model," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Marseille, France.
Krane, M., Barry, M., and Wei, T. (**2007**). "Unsteady behavior of flow in a scaled-up vocal folds model," J. Acoust. Soc. Am. **122**, 3659–3670.
Kucinschi, B., Scherer, R., Witt, K. D., and Ng, T. (**2006**). "An experimental analysis of the pressures and flows within a driven mechanical model of phonation," J. Acoust. Soc. Am. **119**, 3011–3021.
Mongeau, L., Franchek, N., Coker, C., and Kubil, R. (**1997**). "Characteristics of a pulsating jet through a small modulated orifice, with application to voice production," J. Acoust. Soc. Am. **102**, 1121–1133.
Neubauer, J., Miraghaire, R., and Berry, D. (**2004**). "Dynamics of coherent flow structures of a pulsating unsteady glottal jet in human phonation," in American Physical Society 57th Annual Meeting of the Division of Fluid Dynamics, Seattle, WA.
Neubauer, J., Zhang, Z., Miraghaire, R., and Berry, D. (**2007**). "Coherent structures of the near field flow in a selfoscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.
Pelorson, X., Hirschberg, A., Hassel, R., and Wijnands, A. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.
Scherer, R., Titze, I., and Curtis, J. (**1983**). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," J. Acoust. Soc. Am. **73**, 668–676.
Sulter, A., and Wit, H. (**1996**). "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age," J. Acoust. Soc. Am. **100**, 3360–3373.
Triep, M., Brücker, C., and Schröder, W. (**2005**). "High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245.
Van den Berg, J. (**1958**). "Myoelastic-aerodynamic theory of voice production," J. Speech Hear. Res. **1**, 227–244.
White, J. (**1992**). "Evaluation of the pseudoinfinite length tube as used to collect glottal volume velocity waveforms," School of Communications, Northwestern University, Evanston, IL.
Zhang, Z., Mongeau, L., and Frankel, S. (**2002a**). "Broadband sound generation by confined stationary jets through circular and glottis-shaped orifices," in Proceedings of the International Conference on Voice Physiology and Biomechanics, Denver, CO.
Zhang, Z., Mongeau, L., and Frankel, S. (**2002b**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.
Zhang, Z., Mongeau, L., Frankel, S., and Thomson, S. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**, 1720–1728.

A high-speed laryngoscopic investigation of aryepiglottic trilling[a)]

Scott R. Moisik and John H. Esling
Department of Linguistics, University of Victoria, Victoria, British Columbia V8W 3P4, Canada

Lise Crevier-Buchman
Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne-Nouvelle, 19 rue des Bernardins, 75005 Paris, France

(Received 19 December 2008; revised 17 November 2009; accepted 30 November 2009)

Six aryepiglottic trills with varied laryngeal parameters were recorded using high-speed laryngoscopy to investigate the nature of the oscillatory behavior of the upper margin of the epilaryngeal tube. Image analysis techniques were applied to extract data about the patterns of aryepiglottic fold oscillation, with a focus on the oscillatory frequencies of the folds. The acoustic impact of aryepiglottic trilling is also considered, along with possible interactions between the aryepiglottic vibration and vocal fold vibration during the voiced trill. Overall, aryepiglottic trilling is deemed to be correctly labeled as a trill in phonetic terms, while also acting as a means to alter the quality of voicing to be auditorily harsh. In terms of its characterization, aryepiglottic vibration is considerably irregular, but it shows indications of contributing quasi-harmonic excitation of the vocal tract, particularly noticeable under conditions of glottal voicelessness. Aryepiglottic vibrations appear to be largely independent of glottal vibration in terms of oscillatory frequency but can be increased in frequency by increasing overall laryngeal constriction. There is evidence that aryepiglottic vibration induces an alternating vocal fold vibration pattern. It is concluded that aryepiglottic trilling, like ventricular phonation, should be regarded as a complex, if highly irregular, sound source. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3299203]

PACS number(s): 43.70.Gr, 43.70.Fq, 43.70.Jt, 43.70.Bk [AL] Pages: 1548–1558

I. INTRODUCTION

There are numerous structures in the vocal tract apart from the vocal folds that can generate phonatory oscillations when subject to appropriate aerodynamic conditions: the esophagic sphincter, ventricular folds, and aryepiglottic folds are some examples (Catford, 1968; Lindestad *et al.*, 2001; Sakakibara *et al.*, 2004). The use of these structures in generating a voice source has been observed for pathological cases, where vocal fold phonation is not possible or done inefficiently (Fröhlich *et al.*, 2000; Schwarz, 2007). The aryepiglottic folds are unique among these structures because they roughly parallel the vocal folds in their ability to produce both articulations and phonatory voicing sources. The aryepiglottic folds contribute a very distinct category of phonation, typically referred to as growl or harsh voice. Singers such as Louis Armstrong, Louis Prima, Koko Taylor, and Tom Waits readily come to mind as famously exploiting aryepiglottic vibration to produce their unique singing styles, illustrating the application of the sound as a voice quality. Esling (1996) and Edmondson and Esling (2006) argued that the aryepiglottic folds play a central role in forming the primary constriction in the pharynx associated with pharyngeal articulations in general. More specifically, these researchers re-attribute the former categories of epiglottal and pharyngeal to the (active) aryepiglottic and the (passive) epiglottal place of articulation, which forms the locus of pharyngeal constriction, achieved supplementarily by retraction of the tongue (and epiglottis) and by raising of the larynx. Their laryngoscopic evidence shows that numerous languages employ the aryepiglottic folds in a variety of ways to produce linguistic contrasts (Carlson *et al.*, 2004; Edmondson *et al.*, 2005; Esling and Edmondson, 2002; Esling *et al.*, 2005). Articulatory possibilities include stop, fricative, approximant, trill, and possibly even tap.[1] The aryepiglottic sphincter is shown to play a role in the production of creaky[2] and harsh voice, which also can exhibit aryepiglottic trilling/vibration. Additionally, languages containing pharyngeal fricatives and approximants can use aryepiglottic vibration as a means of enhancing the distinctiveness of these sounds (Esling, 2005).

The aryepiglottic folds are anatomically situated at the top of the epilaryngeal tube (also referred to as the laryngeal vestibule; see Fig. 1). The tube can be reduced in volume and ultimately form a protective hermetic seal by means of laryngeal constriction (in the sense of Gauffin, 1977; also see Fink, 1975, pp. 85–102). Titze (2008) argued that this tube plays an important role in coupling the glottal source to the rest of the vocal tract, specifically when the epilaryngeal tube is reduced in cross-sectional area through laryngeal constriction (less than or equal to 0.5 cm^2). Two interactions between the vocal folds and the epilaryngeal tube are identified: First, flow pulse skewing can energize glottal harmonics below the first resonance of the vocal tract and, second, vocal

[a)]This paper is derived from research originally presented at the 6th ICVPB, Tampere, Finland, 6–9 August 2008.

 0001-4966/2010/127(3)/1548/11/$25.00

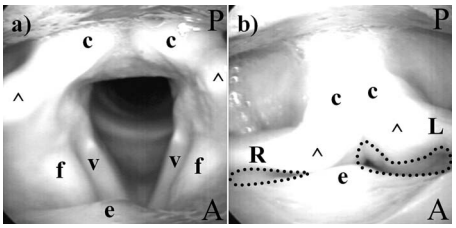


FIG. 1. A superior view of the epilaryngeal tube during the video of a voiced aryepiglottic trill at 100 Hz. Image (a) shows the tube in its unconstricted state (frame 19), and image (b) shows the tube in its constricted state (frame 396). Dotted-line=aryepiglottic aperture, c=corniculate tubercle, ^ =cuneiform tubercle, e=epiglottic tubercle, f=ventricular (false) fold, v =vocal fold, R=right aryepiglottic fold, L=left aryepiglottic fold, A =anterior (front), and P=posterior (back).

fold oscillatory efficiency (the ratio of radiated acoustic power to aerodynamic power) can be enhanced. Using computational models, both of these effects are evinced to be augmented under conditions of reduced epilaryngeal tube area at the expense of destabilizing the oscillation of the vocal folds. Additionally, the epilaryngeal tube also contributes to the resonances of the vocal tract. With respect to the acoustics of the singing voice, Sundberg (1977) claimed that the epilaryngeal tube (or, in his terms, the outlet of the larynx) becomes independent of the vocal tract and produces its own resonance when the cross-sectional area of the epilaryngeal tube is one-sixth the area of the pharynx. Titze and Story (1997) further argued that the epilaryngeal tube is most suitably characterized as a quarter-wave resonator with its first resonance between 2500 and 3000 Hz (also known as the singer's formant). Under voiceless conditions, a turbulent noise source can be produced from an air jet impinging on the posterior surface of the epiglottis and the medial surfaces of the ventricular folds. Stevens (1998, p. 430) estimated that as much as two-thirds of the acoustic energy in glottal fricatives [h] is produced at the surface of the epiglottis. Stevens' (1998) analysis assumes that the pharyngeal airway is not constricted, which presumably means that the epilaryngeal tube is also considered to be unconstricted (i.e., unreduced in cross-sectional area). It is suspected, however, that increased constriction of this tube would also yield an effect on the character of laryngeal-pharyngeal noise sources, which characterize the voiceless pharyngeal fricative /ħ/. For Arabic /ħ/, Butcher and Ahmad (1987) reported high intensity frication noise values with a lower limit between 500 and 1500 Hz and distinct concentrations of noise energy in formant regions associated with the vowel context.

Overall, laryngeal constriction causes the epilaryngeal tube to change the acoustic and aerodynamic properties of the vocal tract. In addition to these effects, the oscillatory capacity of the epilaryngeal tube represents a further degree of complexity of the laryngeal system as a whole. Specifically, it is posited that aryepiglottic trilling can be regarded as a vocal tract source, one that can interact with the glottal source but that is also independent from it. The purpose of this paper is to adduce evidence that supports this characterization and provides an illustration of the oscillatory nature of aryepiglottic trilling in general. The assertion that the aryepiglottic folds are source generators will be supported by comparing aryepiglottic aperture data and kymographic traces with the corresponding acoustic waveforms and electroglottograph (EGG) signals for several aryepiglottic trills with various glottal settings (i.e., pitch and the presence or absence of voicing). The appropriateness of the phonetic label *trill* will also be a central point of discussion, and the decision to use the label is motivated by the data brought forth in this study.

II. HIGH-SPEED LARYNGOSCOPIC DATA

High-speed laryngoscopic videos of six different aryepiglottic trills were studied. All of the trilling sequences were performed by Esling, a 60-year-old male phonetician, who was trained in the Edinburgh tradition of phonetics. Two sets of aryepiglottic trills were performed with an [ə] vocalic target: one voiced and one voiceless. In the voiced set, glottal pitch was targeted at 100 and 200 Hz. In the voiceless set, the subject produced the trills by emulating the voiced configurations required for 100 and 200 Hz while keeping the vocal folds abducted (so that the sound produced would be voiceless). An additional parameter for both voiced and voiceless trills was explored: degree of laryngeal constriction. These are called "tense" trills (cf. Heselwood, 2007), although no claims as to the exact degree of muscular tension are made; rather, the term reflects the subject's own interpretation of his performance, expressing the extra effort or tension in producing these sounds. The performances were facilitated by the use of pre-recorded target sounds, which the subject listened to using headphones during the video recording session.

The system used to capture the trilling sequences was an SL Kamera 500 connected from a rigid oral endoscope to a SpeedCam Lite interface (Weinberger, Erlangen, Germany), which was set to a frame rate of 500 Hz (one frame for every 2 ms) with a resolution of 262×256 pixels. Recording was done at the Sorbonne-Nouvelle Paris III/CNRS-LPP-UMR 7018 research site at l'Hôpital Européen Georges Pompidou, Paris. The average sequence length was 1.945 s (or around 972 frames). Throughout the recording session, the audio signal of the trills was obtained using a head mounted AKG C410 microphone, positioned at 4 cm from the lips, with an angle of 45° from sagittal plane, sampled at 44 100 Hz, 16-bit resolution. The EGG signal was obtained using a Glottal Enterprises EG2-PCX electroglottograph; the surface electrodes were placed externally on each thyroid lamina, which was constant for all trill performances. The audio and video signals were digitized using ANAVOX (custom software; Vannier-Photelec, Antony, France). A time-to-live signal was used to ensure synchronization of the audio with the video signal. The videos were digitized frame by frame in uncompressed format and later stored as uncompressed TIFF files for analysis in MATLAB.

III. AUTOMATED IMAGE ANALYSIS

An automated image analysis technique was employed to analyze the high-speed laryngoscopic videos. This involved assessing the percent of aperture formation for each of the aryepiglottic folds during oscillation. These data are

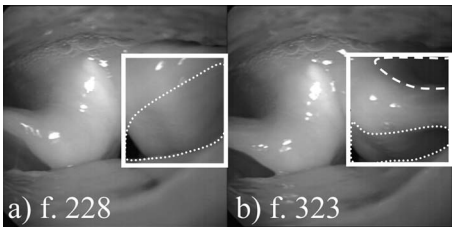


FIG. 2. Non-tense voiceless aryepiglottic trilling. The left fold undergoes extreme displacement, making ROI placement (solid rectangle) over the aperture (dotted-line) difficult. In (a) the left aperture is wide and requires a large ROI in frame 228 of the sequence. The same aperture in (b), however, at frame 323 encloses area over the piriform fossa (dashed-line) in addition to the aryepiglottic aperture.

accompanied by kymographic images and EGG signal for comparison. The following sections discuss this analysis in detail.

A. Aperture area analysis

The aperture analysis involved the identification of regions of interest (ROIs) across the sequence of video frames for both of the aryepiglottic apertures (defined by the subject's right and left aryepiglottic folds; see Fig. 1) and then processing these regions to yield an image that could be measured using MATLAB's image processing functions, e.g., `bwarea()`. Two problematic circumstances arose that required special ROI placement for the left aryepiglottic aperture. For the normal and tense voiceless aryepiglottic trills at 100 Hz, the left aryepiglottic aperture fluctuated so much that it was difficult to select a ROI that maximally excluded non-aperture related areas (primarily the piriform fossa). This is illustrated in Fig. 2(a), where a ROI (solid line; not used in the actual analysis) is placed over an area large enough to enclose an extreme displacement of the left aryepiglottic fold (dotted-line). Later in the trilling sequence [Fig. 2(b)], the same ROI encloses a large portion of the space associated with the piriform fossa (dashed-line), in addition the aryepiglottic aperture itself. Another problem was observed for the voiced aryepiglottic trill at 200 Hz, where the epiglottis occluded the view of the left aryepiglottic aperture entirely, as seen in Fig. 3. In both of these extenuating cases, it was decided that oscillation information for the left aryepiglottic aperture could be still obtained by placing the ROIs over the left piriform fossa instead. Thus, for these cases, the ROIs used in the analysis monitor the position of the posterior surface of the left aryepiglottic fold as it occludes the view of the left piriform fossa. The resultant signal was inverted to match the phase of the actual aperture, which was assumed to be opposite to the signal (i.e., the observed piriform fossa area enlarges when the left aryepiglottic aperture diminishes). The drawback of this approach is that any subtle, minor vibrations (i.e., surface mucosal waves) only visible on the inner portion of the fold are lost; this was deemed an acceptable sacrifice so that at least major pulse information could be collected for the offending fold.

ROI processing is illustrated in Fig. 4; the leftmost image shows frame 236 of the voiced aryepiglottic trill at 100 Hz. In this frame, both the right and left ROIs are outlined. The processing steps used for all ROIs are demonstrated for

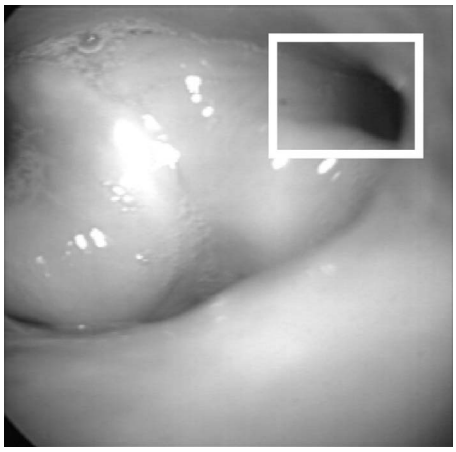


FIG. 3. Frame 166 of the "200 Hz" voiceless aryepiglottic trill. The left aperture is barely visible, but oscillation information was obtained by tracking a ROI (solid rectangle) placed over the left piriform sinus (upper right hand corner) and then inverting the signal. The right aperture is visible as a dark sliver (bottom left hand corner).

the right aperture in the images on the right [Figs. 4(a)–4(c)]. The ROIs were first extracted [Fig. 4(a)] from their corresponding frame and then inverted [Fig. 4(b)]. Before area measurement can be made, the image needs to be in binary representation [Fig. 4(c)]. This step requires a black-white luminance threshold value to be selected. The threshold values were determined automatically (prior to the actual measurement routine) by calculating the average grayscale value for each ROI from its grayscale histogram using the `imhist()` function in MATLAB. An average of all of the mean grayscale values for all ROIs in a given sequence was obtained and converted to a luminance threshold value by dividing by 255 (the number of values in an 8-bit grayscale image). Aperture data are expressed as a percentage of the maximum aperture area value recorded for a particular sequence and fold. Exact values for all parameters of the image analysis are presented in Table I. The overall process was automated using simple MATLAB scripts.

As a complement to the aperture data, kymographic images (e.g., Švec and Schutte, 1996) were created for both the left and right aryepiglottic apertures. A line of pixels (the kymographic line) was selected for each aperture, which was

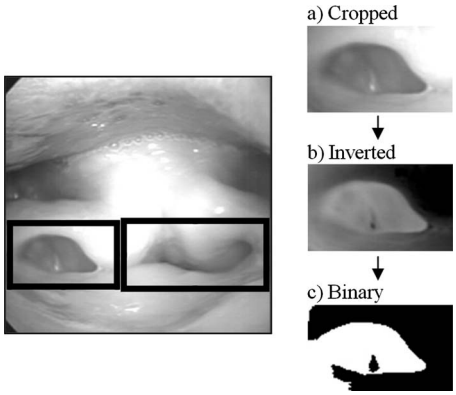


FIG. 4. Illustration of ROI definition over the left (larger rectangle) and right (smaller rectangle) aryepiglottic apertures and image processing sequence to arrive at a binary image that can be measured for aperture area.

TABLE I. A listing of values for the aperture analysis and related details of the high-speed videos of voiced (vd.) and voiceless (vl.) aryepiglottic trilling. Height and width dimensions are in pixels. CC analysis=cross-correlation analysis between aperture and kymographic data. "Yes" in the "inverted data" row indicates that the aperture data were inverted (see text for explanation).

	100 Hz		200 Hz		100 Hz tense	
Sequence	vd.	vl.	vd.	vl.	vd.	vl.
Start frame	467	191	120	443	42	1041
End frame	1441	2046	335	1555	1047	1861
No. of frames	974	1855	215	1112	1005	676
Duration (s)	1.948	3.710	0.430	2.224	2.010	1.352

Aperture	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Threshold	0.10	0.09	0.54	0.73	0.46	0.51	0.24	0.21	0.42	0.51	0.38	0.67
ROI width	101	131	100	102	53	84	43	151	45	99	115	93
ROI height	57	73	68	68	74	129	49	86	44	35	85	99
CC analysis	0.976	0.978	0.885	0.864	0.921	0.907	0.940	0.937	0.946	0.980	0.997	0.939
Inverted data	No	No	No	Yes	No	Yes	No	No	No	No	No	Yes

different in every video. The lines were selected such that they were perpendicular to the visible section of the fold (at rest) and bisected it. In some cases, due to camera position (as in Fig. 2), only a small portion of the fold was visible. In this case, the line was placed such that it passed through the point on the visible fold that exhibited the greatest displacement. Line length was decided by viewing the entire video and determining the minimal length required to capture as much of the aperture oscillation as possible while excluding non-aperture related image information. This process was complicated by the fact that in several videos the position of the larynx changes slightly through the performance of the aryepiglottic trill, which means that the image of the larynx migrates somewhat. Ultimately, the lines were only used if all of the researchers agreed that they represented the optimal choice given the video data.

The data obtained from the kymographic images were then converted into an oscillatory signal simply by summing the grayscale values for each line in the kymographic image. In the case that a kymographic image included non-aperture data (e.g., an image of the piriform fossa due to migration of the image of the larynx), these areas were masked by setting the pixel values to zero. The kymographic "signal" was then compared to the aperture data using cross-correlation analysis to quantify how similar the two signals were. Successful matching of the signals was defined to be indicated by a normalized zero-lag component of the cross-correlation analysis above 0.7071.[3] None of the signals obtained in this method resulted in a cross-correlation less than 0.7071; values of the analysis are presented in Table I.

The final technique used to analyze the aperture data was peak detection analysis, using a MATLAB script developed by O'Haver (2006), which allowed the average frequency and jitter of the EGG and aperture signals to be calculated. A peak detection analysis window of 50 ms was used for all signals; in the case of the aperture data, this window typically included two to five aryepiglottic pulses. There was a 25 ms overlap interval between analysis windows. The frequency of the signal for each analysis window was determined to be the inverse of the average peak-to-peak (the signal was always positive) interval time; the frequency for the entire signal was the average of all analysis window results. Jitter was calculated by taking the standard deviation of all peak-to-peak durations in the signal.

Sections III A 1–III A 4 present a selection of data from the analysis described above that serve to characterize the nature of aryepiglottic trilling for our subject. Sections III A 1 and III A 2 examine the basic aryepiglottic trill in voiced and voiceless contexts; in both cases a contiguous set of 25 video frames is presented alongside an array of synchronized plots of the aryepiglottic aperture data, the kymographic images, and the audio signal. Section III A 3 documents the tense voiceless aryepiglottic trill. Section III A 4 presents data for all trills including acoustic spectra and Table II, which contains root mean square (rms) intensity values and peak detection analysis values.

1. Voiced aryepiglottic trill

Twenty-five frames of the voiced aryepiglottic trill produced at a glottal frequency of 100 Hz are presented in Fig. 5. The lines used to create the corresponding kymographic images have been superimposed on the first frame. This set corresponds exactly to a time period between 0.800 and 0.850 s in the data plots displayed in Fig. 6. In this time period, two right aryepiglottic fold pulses can be seen; the left fold also exhibits two major pulses, each with an additional minor pulse evident both in the kymographic image and the aperture signal. Throughout the entire trill, the right aryepiglottic fold oscillates semi-periodically; it typically makes complete contact with the surface of the epiglottis. Additional pulses tend to occur when complete contact is not made (e.g., at 0.860, 1.030, and 1.170 s). The gross movements of the left aryepiglottic fold appear to be largely in phase with the right fold during the section shown in the plots and have roughly the same frequency (see Table II); however, very often in this sequence the left fold oscillates with major and minor pulses, particularly between 0.800 and 1.100 s; after this point the pulsing diminishes in magnitude and the alternating major-minor pattern is not discernible. By

TABLE II. Intensity and frequency values for the audio signals of the voiced (vd.) and voiceless (vl.) aryepiglottic trills. rms intensity was obtained for the 200 ms region between the listed analysis frames. Frequency data are averaged across the entire trilling sequence.

	100 Hz		200 Hz		100 Hz tense	
Sequence	vd.	vl.	vd.	vl.	vd.	vl.
Analysis frames	967–1067	341–441	221–321	1193–1393	947–1047	1191–1291
rms intensity (dB)	−9.0309	−17.4958	−7.6700	−17.8252	−9.7551	−10.8040
Actual F_0 (Hz)	104.0	⋯	200.3	⋯	110.6	⋯
Glottal jitter (ms)	0.32	⋯	0.53	⋯	0.07	⋯

Aperture	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Average frequency (Hz)	65.9	82.4	67.4	42.4	64.7	69.8	55.8	100.4	92.0	100.8	73.8	71.6
Jitter (ms)	5.8	6.7	3.6	8.9	3.4	7.4	5.9	2.7	2.6	1.6	5.1	3.0

observing the videos themselves, the minor pulse of the left aryepiglottic fold appears to be a mucosal wave traveling on the aryepiglottic fold in the cranial (cephalad) direction. Comparison of the EGG signal with the audio and aperture data signals indicates that the aryepiglottic pulse actually corresponds to perturbations of the vibratory pattern of the vocal folds. A double peaked (or plateau-peak type) glottal pulse tends to occur every time the right aryepiglottic aperture is open (e.g., the pulse immediately after 1.000 s). The increase in noise in the acoustic signal at this point indicates that turbulent air is escaping from the glottis, which is open for a prolonged duration during the open phase of the right aryepiglottic pulse. When the right aryepiglottic aperture is closed, the glottis tends to open and close very rapidly. The left aryepiglottic aperture does not appear to have any discernible impact on the vocal folds, although its continually changing size is likely to impact supraglottal pressure.

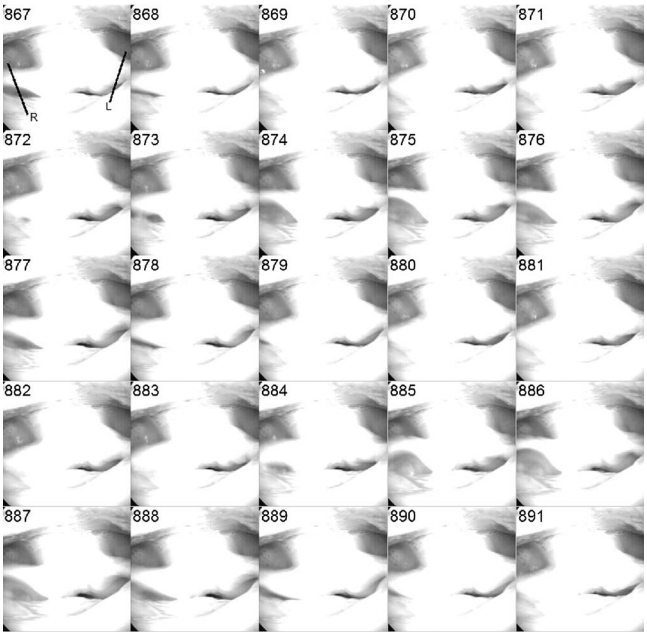


FIG. 5. Twenty-five frames of voiced aryepiglottic trilling (100 Hz glottal pitch, 25 frames, 50 ms). First frame (867) shows right and left lines used to extract kymographic images. Corresponds to Fig. 6 between 0.800 and 0.850 s.

2. Voiceless aryepiglottic trill

The first twenty-five frames and corresponding aperture and kymographic data of the voiceless aryepiglottic trill produced at an emulated 100 Hz can be seen in Figs. 7 and 8 (see Table II for frequency and intensity data). It is crucial to note that there was no actual glottal pulse, which is corroborated by the (essentially) flat EGG signal. Although the audio signal is considerably noisy (indicative of increased airflow due to glottal voicelessness), a distinctive pulse pattern can be seen, which roughly corresponds to the signals in the aperture data. Unlike the voiced aryepiglottic trill at 100 Hz, this trill exhibits massive left aryepiglottic fold displacement, as seen in the video frames (Fig. 7). Particularly interesting is the behavior of the left cuneiform cartilage, which can be seen to oscillate like a rubbery/highly elastic rod.

3. The tense voiceless aryepiglottic trill

An alternative voiceless aryepiglottic trill was performed by our subject; this sequence is considered to be a tense variant of the voiceless trill, although the degree of tension is not strictly quantifiable. Despite the vagary of its characterization, the data in Fig. 9 indicate that this trill merits consideration. It should be noted again that the sound is voiceless, confirmed by the relatively flat EGG signal. The right aryepiglottic pulse pattern shows alternating major pulse and minor pulses: The acoustic effect associated with the major pulse can be clearly seen in the audio signal as a distinctive pressure spike, particularly in the section between 0.450 and 0.700 s. The negative part of this spike appears to coincide with the opening phase of the right aryepiglottic pulse, while the positive part occurs when the aperture is more fully open. Between 0.300 and 0.500 s, the left aryepiglottic fold tends to exhibit the major-minor pattern. The minor pulses must occur for the entire fold (not just the interior surface) because the left aryepiglottic aperture data, in this case, are based on measurements above the piriform fossa. Between 0.500 and 0.700 s, the left aryepiglottic fold begins to exhibit a more steady oscillatory pattern with approximately equal intensity pulses.

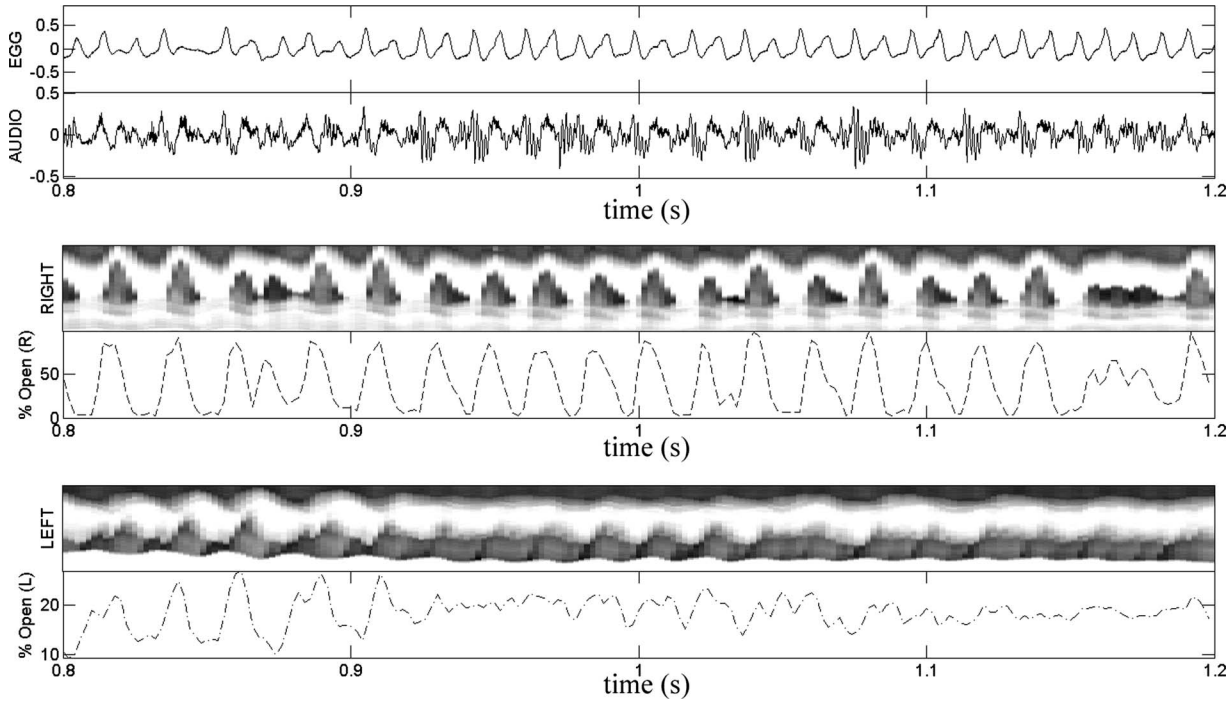


FIG. 6. Voiced aryepiglottic trilling (100 Hz glottal pitch): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

4. Aryepiglottic trill frequency data

Figure 10 contains spectra for all six of the trills. All of the spectra were obtained using a fast Fourier transform calculated over a 100 ms section of relative signal stability, using a 20 Hz Gaussian window. As the signals are highly noisy, the lower harmonic content was best visualized by limiting the spectrum display from 0 to 500 Hz. The voiced aryepiglottic trills all have fairly well defined and intense glottal fundamentals, although the higher harmonics are not always clearly defined. The aryepiglottic pulses arguably explain the presence of nearly harmonic energy around the 50 Hz mark (in the 100 and 200 Hz voiced trills), although the noise in the signal and the jitter of the aryepiglottic pulse likely smear the pulse energy over a broader frequency range than the glottal pulse. The voiceless trills show harmonic spikes at values that correspond well with the values measured in the peak detection analysis (see Table II). The tense voiceless aryepiglottic trill is the exception to this where the spectral energy is exceedingly noisy; the peak at 70 Hz might correspond with the measured frequency of the folds, but the stronger peak at ~30 Hz is possibly attributable to the major pulse of the right aryepiglottic fold that characterizes this trill. As mentioned in Sec. III A 3, the fundamental frequency of 73.8 Hz for the right fold during this trill was estimated based on both the major and minor pulses. If the minor pulses are ignored, then the mean frequency for the right fold in the tense trill is 31.2 Hz, which agrees with the spectrum of this sound.

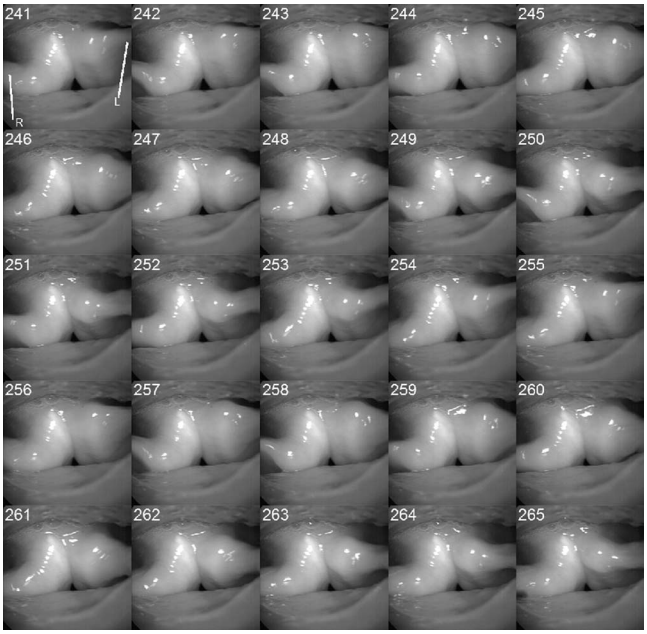


FIG. 7. Twenty-five frames of voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration, 25 frames, 50 ms). First frame (241) shows right and left lines used to extract kymographic images. Corresponds to Fig. 8 between 0.100 and 0.150 s.

Table II provides a listing of all measurements pertaining to the intensity of the overall signal and frequency of the vocal and aryepiglottic fold pulses. In all voiced cases, the average frequency of the glottal pulse (calculated from the EGG signal) corresponds well with the desired target pitches (100 and 200 Hz). In the case of our subject, the right aryepiglottic fold exhibits a frequency range between 55.8 and 92.0 Hz. The subject's left aryepiglottic fold was more difficult to measure (as discussed above in Sec. III A); it ranged between 42.4 and 100.8 Hz. Although not entirely consistent, the tense trills can be said to increase the pulse frequency for both of our subject's aryepiglottic folds by roughly 10–40 Hz.

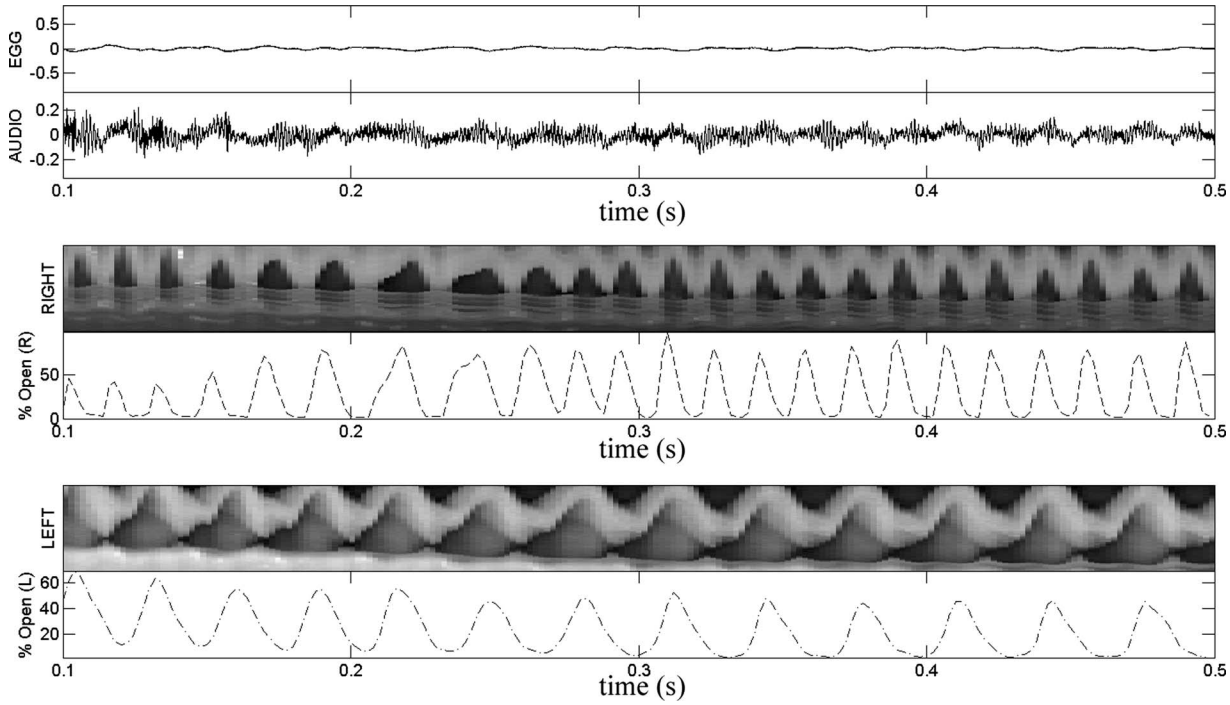


FIG. 8. Voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

The rms intensity values for the signals reveal a distinction between the non-tense voiced and voiceless trills: The voiced trills were approximately 8–10 dB more intense than the voiceless ones; however, the tense trills (voiced and voiceless) were of similar intensity. The acoustic spectra (Fig. 10) also show that the voiced trills are generally more intense than the voiceless ones. This difference is likely due to increased turbulent losses during the voiceless trills. The aryepiglottic folds are evidently not as efficient at contouring the pulmonic airstream into a pulsatile flow as the vocal folds are.

IV. DISCUSSION

Even within the performance of a single individual, the variability of aryepiglottic fold trilling behavior is mani-

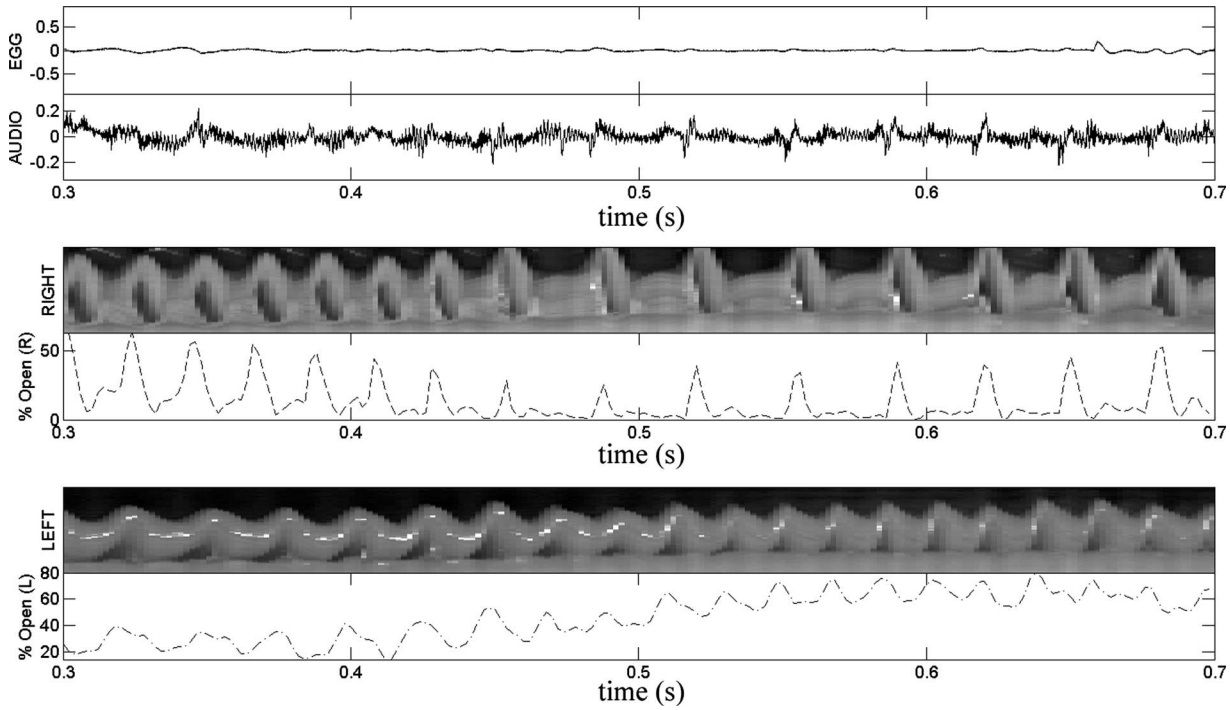


FIG. 9. Tense voiceless aryepiglottic trilling (emulated 100 Hz glottal pitch configuration): EGG signal and oscillogram (top), right aryepiglottic fold kymograph and aperture data (middle), and left aryepiglottic kymograph and aperture data (bottom).

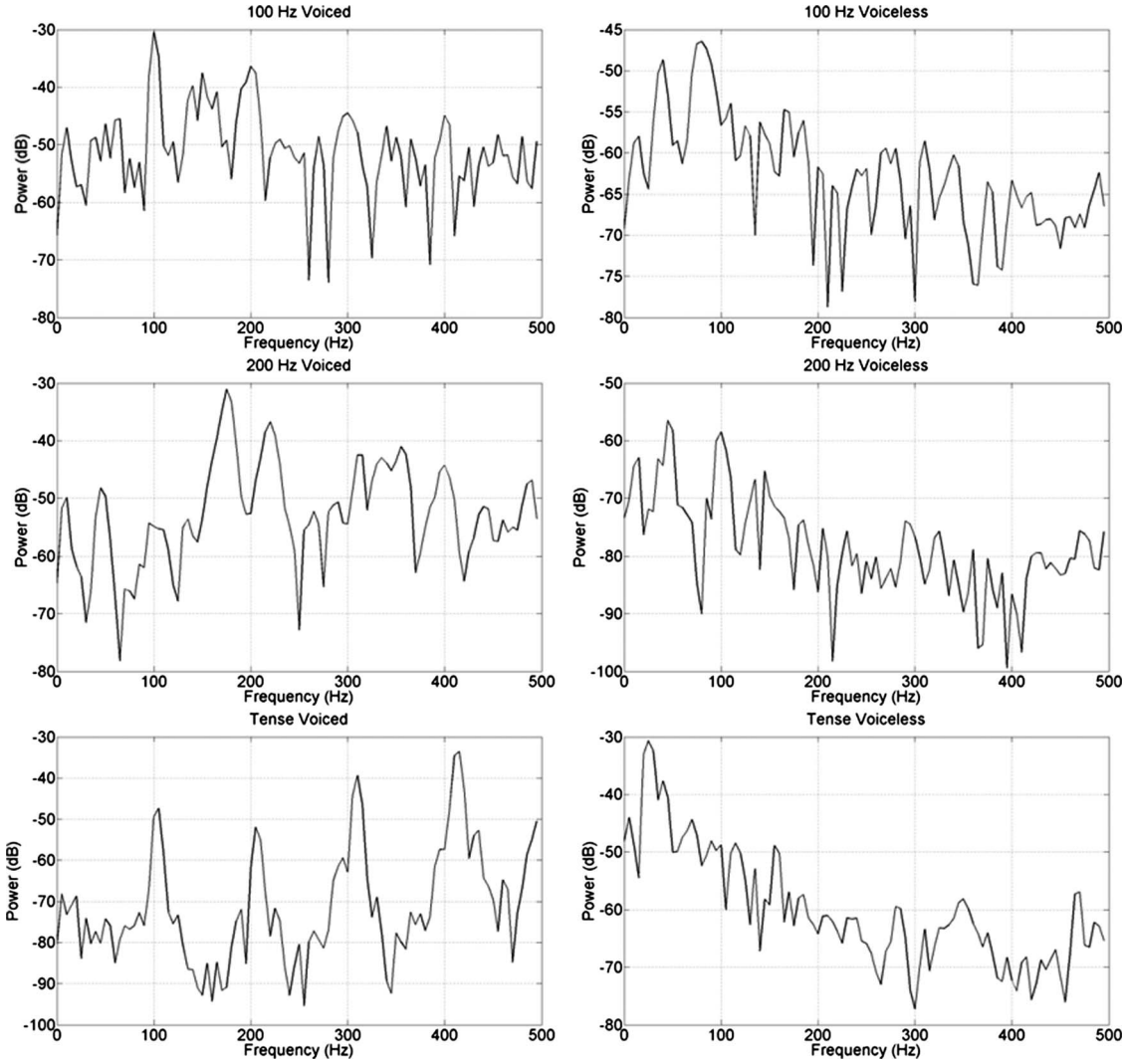


FIG. 10. Spectra of all trills (steady 100 ms interval used). Left columns are voiced trills (100 Hz, ~200 Hz, 100 Hz tense), and right columns are voiceless trills (emulated 100 Hz, 200 Hz, tense).

fested. Three major parameters were employed to study aryepiglottic trilling: the presence or absence of vocal fold vibration, glottal fundamental frequency, and the overall extent of laryngeal constriction. Although the latter parameter is only described qualitatively, it must be possible to vary constriction due to the fact that the aryepiglottic sphincter varies in configuration from being fully open to being fully closed, which is well documented in the languages of the world (e.g., Esling, 1996; Edmondson and Esling, 2006). The 100 Hz set of aryepiglottic trills exhibits similar oscillatory behavior despite one trill being voiced and the other being voiceless. The reader is reminded that the left aryepiglottic fold frequency value measured for the voiceless trill (see Table II) is only indicative of the major pulse, as the actual aperture could not be measured. Thus, it is possible that the value reported is half of its actual value, which depends on the presence of the minor pulses, which can only be observed on the interior of the fold. The 200 Hz set revealed two notable changes in the oscillatory frequency. The voiced trill of this set behaves similarly to its 100 Hz counterpart, with the exception that the left aryepiglottic fold exhibited slower oscillations. The voiceless trill of this set, however, was remarkably different: The right aryepiglottic fold decreases in frequency, while the left aryepiglottic fold considerably increases in frequency. The aperture data from this trill (not presented in this article) show that the left aryepiglottic fold oscillates in a particularly irregular way and with variable intensity; it does, however, consistently make full contact with the base of the epiglottis. It is noteworthy that, despite a doubling of glottal pitch, the aryepiglottic folds themselves do not double in frequency. Finally, the tense trills generally exhibited increased oscillatory frequency for both of the aryepiglottic folds (except for the left fold in the voiceless trill, which dropped in frequency slightly compared to the "non-tense" 100 Hz trills). For the voiced trill of this set, it is speculated, based on the frequency values recorded in Table II, that there is greater coupling between the glottal and aryepiglottal systems. With increased tension, the aryepiglottic folds could be more sensitive to the driving pressure generated by the glottal pulse, which is reflected in the fact that aryepiglottic fold frequencies almost match that of the glottal pulse for this trill.

There are two major theoretical questions of interest for this study. The first is whether the aryepiglottic aperture can

be considered a sound source. Both oscillographic and spectral evidence adduced in this study suggest that the aryepiglottic folds contribute to the harmonic spectrum of the signal. This is particularly evident for the voiceless trills where the EGG evidence indicates that the (near) periodic signal cannot be emanating from the glottis. The high degree of jitter in aryepiglottic fold oscillation, however, makes it difficult to clearly assess the exact harmonic content they are generating, particularly since both folds can oscillate at different frequencies, with major and minor pulsations, and in and out of phase with each other. Furthermore, both folds show open oscillations (where the aryepiglottic fold does not make complete contact with the epiglottis) and close oscillations, where collision with the posterior surface of the epiglottis might also generate acoustic energy. From the aperture analysis data and spectra, the frequency of the aryepiglottic folds was determined to tend toward a range between 50 and 100 Hz in the case of our subject.

While the voiced aryepiglottic trill poses a considerable challenge in determining what the exact contribution of the aryepiglottic fold oscillation to the overall laryngeal sound source is, the voiceless aryepiglottic trill provides some evidence that the folds do indeed contribute a pulsatile excitation to the vocal tract, suggesting that the folds can act as an alternative laryngeal source. For example, when the vocal folds are entirely absent (i.e., due to surgical resection of the vocal folds; Fröhlich *et al.*, 1998; Fröhlich *et al.*, 2000; Schwarz, 2007), the aryepiglottic folds function as the voicing source. The aryepiglottic source, however, appears far less efficient at converting the pulmonic airstream into acoustic energy than the vocal folds: The voiceless aryepiglottic trills were consistently less intense than the voiced trill, and the oscillograms and spectra both indicate that the acoustic signals these trills generate are highly noisy. This could be attributed to the fact that much of the air is escaping non-uniformly through one or both of the aryepiglottic apertures, in addition to the medial opening between the arytenoid cartilages, which is never fully sealed during any of the trills. Undoubtedly this amounts to considerable turbulent loss of aerodynamic power. The exact acoustic and aerodynamic nature of the turbulence generated by the aryepiglottic constriction is crucial to understanding pharyngeal fricatives and requires further study.

Throughout this study, the term trill has been used as a phonetic characterization of the oscillatory behavior of the aryepiglottic folds. This investigation reveals that this label of manner of articulation is appropriate for a number of reasons. While aryepiglottic trilling is unique due to the potential for two structures (the left and right aryepiglottic folds) to vibrate simultaneously and yet with considerably different characters (in terms of frequency, phase, and displacements), they arguably exhibit the same physiological conditions and restrictions of the other trilling structures, such as the tongue tip, uvula, and lips. Anatomically, trilling is claimed to be optimized when the tissue undergoing the trill is well supported by adjacent structures and forms a "funnel-like constriction, with the articulator opening along the direction of the airstream" (Barry, 1997, p. 36). Fundamentally, this configuration takes advantage of the Bernoulli effect when a sufficient airstream is generated (Catford, 2001, p. 173). When the aryepiglottic folds are in their constricted configuration, their opening is indeed in the outward direction of the glottal airstream. The fact that all of the trills in this study were produced with considerable antero-posterior approximation of the epilaryngeal tube and with pharyngeal narrowing reflects the requirement for the aryepiglottic folds to be supported by adjacent structures. No aryepiglottic trilling has been attested, as far as the authors are aware, without these conditions holding (for example, in an entirely unconstricted state as occurs for modal voice). Furthermore, regardless of whether one or both of the folds are oscillating, the end acoustic product is uniformly describable as a trilling sound produced at the aryepiglottic place of articulation.

Aryepiglottic oscillations are also trill-like in the sense that there are no real intrinsic or specialized mechanisms for aryepiglottic pitch control. Vocal fold vibration can be controlled intrinsically with the contraction of the thyroarytenoid muscles and extrinsically relies on the highly specialized cricothyroid mechanism of pitch control. As this study demonstrates, the aryepiglottic folds are considerably limited in terms of their frequency range. The highest frequencies were observed for the tense trills, which indicates that general laryngeal constriction can alter the biomechanical nature of the folds to some degree, perhaps increasing the tension on the folds and resulting in a slightly increased oscillatory frequency (of about 20–40 Hz). Other trills are similar in this regard; they do not permit excessive manipulation of their frequency.

Despite these shared characteristics, the aryepiglottic trill must still be regarded as unique among trills because of its low position in the vocal tract. While it is possible to excite the resonances of the vocal tract using buccal or uvular trills, producing nearly unintelligible speech, aryepiglottic trilling is phonatory in the sense that it can be used as a quality of voicing (described as a category of *harsh voice*), and this function is incorporated into the linguistic systems of several languages, such as !Xóõ (Tuu; Traill, 1986), N|uu (Tuu; Miller *et al.*, 2009), and Jianchuan Bai (Tibeto-Burman; Edmondson and Esling, 2006).

The aryepiglottic trilling documented here might also be compared to the phonation described for various ethnic singing styles of Asia, such as Mongolian throat singing. One of the important conclusions of Lindestad *et al.* (2001) is that a ventricular oscillation half-periodic to the glottal pulse acts to dampen the glottal source on every second glottal pulse. They regarded this mode of phonation as a complex sound source, where the ventricular damping is considered to contribute to a very rich harmonic spectrum that is characteristic of the singing style (which employs overtone enhancement). A similar alternating glottal-damping effect might occur for voiced aryepiglottic trill at 100 Hz (e.g., Fig. 6). In this particular trill, the aryepiglottic occlusion is not complete: Only the right aryepiglottic fold fully closes; the left aryepiglottic aperture, on the other hand, only reduces in area. Thus, the glottal damping in this case is not complete. Rather, once the right aryepiglottic aperture opens, the vocal folds destabilize (exhibiting a double peaked oscillation) and the acoustic signal becomes more noisy. Furthermore, the oscillatory behav-

ior of the aryepiglottic folds is generally more irregular than that of the ventricular folds, which appear to oscillate in sympathy with the vocal folds in Lindestad *et al.*'s (2001) data.

A note must be made about the question of how the acoustic impedances are changed in the vocal tract under aryepiglottic oscillation. As already described, under conditions of static laryngeal constriction, Titze's (2008) *nonlinear source-filter coupling in phonation theory* predicts that the epilaryngeal tube contributes a positive reactance (inertance) to the vocal tract, which can both enhance the efficiency of glottal phonation as well as destabilize it. It is a point for future research to determine how this interaction might change under conditions of aryepiglottic trilling, where the epilaryngeal tube can potentially be separated into two outlet channels (the left and right apertures) that deform over time.

A significant challenge will be the simulation of aryepiglottic trilling in a computational model. One of the purposes of the present research is to contribute to the general understanding of aryepiglottic trilling with the goal of developing a computational simulation of the phenomenon. Much research remains, however, particularly with regard to understanding the acoustic nature of the trill and the aerodynamic conditions under which it occurs. Furthermore, there is little documentation available on the biomechanical properties of the aryepiglottic folds themselves. Computational modeling of the structures would greatly benefit from specific knowledge of their biomechanical characteristics and even their dimensions and mass.

V. CONCLUSION

The purpose of this study has been to understand the oscillatory nature of aryepiglottic trilling. Evidence derived from image analysis of high-speed laryngoscopic videos of six trills with various glottal parameters indicates that the aryepiglottic folds can behave as an acoustic source to the vocal tract under conditions of glottal voicelessness, and perturb the vibration of the vocal folds by triggering alternation in the character of the glottal pulse. On account of this, the voiced aryepiglottic trill should be considered a complex sound source that modulates the quality of voicing and introduces noise as a result of prolonged glottal opening. The voiceless aryepiglottic trills are characterized by being much less intense than the voiced ones, although semi-harmonic excitation of the vocal tract occurs due to the aryepiglottic fold oscillations. Most of the acoustic power for the voiceless trills is converted into noise, which is characteristic of this sound. The aryepiglottic trill is deemed to be appropriately labeled as a trill primarily due its independence of the vocal folds and limited means of pitch control. It was found that increasing the "tension" of the overall laryngeal constriction used in producing the trill could increase the oscillating frequency of the aryepiglottic folds.

Future research must be carried out on other subjects and using linguistic data to determine the strength of these claims. Our next objective will be to analyze aryepiglottic trilling as it occurs in attested words of Iraqi Arabic with the image analysis techniques used for this study. One of the ultimate goals is to create a computational simulation of aryepiglottic trilling. The data collected here will be used as a basis for such a model.

[1]Based on Esling (1996), the International Phonetic Alphabet symbols used to represent these sounds are as follows: stop [ʔ], fricative [ħ], approximant [ʕ], and trills [ʜ ʢ]. The pharyngeal tap [ʡ̆] is also attested.

[2]Lindqvist (1972) also argued that the aryepiglottic sphincter played an essential role in the production of creaky voice.

[3]A value of 1 indicates total positive correlation; in MATLAB this type of normalization is obtained by setting the `xcorr()` option flag to "`coeff`." The limiting value of 0.7071 was selected because it represents the zero-lag component of a cross-correlation analysis of two sine-waves that are exactly 45° out of phase.

Barry, W. (**1997**). "Another [r]-tickle," J. Int. Phonetic Assoc. **27**, 35–45.

Butcher, A., and Ahmad, K. (**1987**). "Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic," Phonetica **44**, 156–172.

Carlson, B. F., Esling, J. H., and Harris, J. G. (**2004**). "A laryngoscopy phonetic study of Nlaka'pamux (Thompson) Salish glottal stop, glottalized resonants, and pharyngeals," in *Studies in Salish Linguistics in Honor of M. Dale Kinkade*, Occasional Papers in Linguistics Vol. **17**, edited by D. B. Gerdts and L. Matthewson (University of Montana Press, Missoula, MT), pp. 58–71.

Catford, J. C. (**1968**). "The articulatory possibilities of man," in *Manual of Phonetics*, edited by B. Malmberg (North-Holland, Amsterdam), pp. 309–333.

Catford, J. C. (**2001**). "On Rs, rhotacism and paleophony," J. Int. Phonetic Assoc. **31**, 171–185.

Edmondson, J. A., and Esling, J. H. (**2006**). "The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopy case studies," Phonology **23**, 157–191.

Edmondson, J. A., Esling, J. H., Harris, J. G., and Huang, T. (**2005**). "A laryngoscopic study of glottal and epiglottal/pharyngeal stop and continuant articulations in Amis—An Austronesian language of Taiwan," Language and Linguistics **6**, 381–396.

Esling, J. H. (**1996**). "Pharyngeal consonants and the aryepiglottic sphincter," J. Int. Phonetic Assoc. **26**, 65–88.

Esling, J. H. (**2005**). "There are no back vowels: The laryngeal articulator model," Can. J. Ling. **50**, 13–44.

Esling, J. H., and Edmondson, J. A. (**2002**). "The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai," in *Phonetics and Its Applications: Festschrift for Jens-Peter Köster on the Occasion of His 60th Birthday*, edited by A. Braun and H. R. Masthoff (Franz Steiner, Stuttgart), pp. 38–51.

Esling, J. H., Fraser, K. E., and Harris, J. G. (**2005**). "Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka)," J. Phonetics **33**, 383–410.

Fink, B. R. (**1975**). *The Human Larynx: A Functional Study* (Raven, New York).

Fröhlich, M., Michaelis, D., and Strube, H. W. (**1998**). "Acoustic 'breathiness measures' in the description of pathologic voices," Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 1998, Seattle, WA)**2**, 937–940.

Fröhlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (**2000**). "Acoustic voice analysis by means of the hoarseness diagram," J. Speech Lang. Hear. Res. **43**, 706–720.

Gauffin, J. (**1977**). "Mechanisms of larynx tube constriction," Phonetica **34**, 307–309.

Heselwood, B. (**2007**). "The 'tight approximant' variant of the Arabic 'ayn," J. Int. Phonetic Assoc. **37**, 1–32.

Lindestad, P. A., Sodersten, M., Merker, B., and Granqvist, S. (**2001**). "Voice source characteristics in Mongolian 'throat singing' studied with high-speed imaging technique, acoustic spectra, and inverse filtering," J. Voice **15**, 78–85.

Lindqvist, J. (**1972**). "A descriptive model of laryngeal articulation in speech," Speech Transm. Lab. Q. Prog. Status Rep., Stockholm, KTH **13**, 1–9.

Miller, A., Brugman, J., Sands, B., Namaseb, L., Exter, M., and Collins, C.

(**2009**). "Differences in airstream and posterior place of articulation among N|uu clicks," J. Int. Phonetic Assoc. **39**, 129–161.
O'Haver, T. (**2006**). "Peak finding and measurement. Custom scripts for the Matlab platform," http://www.wam.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm (Last viewed 6/10/2008).
Sakakibara, K. I., Fuks, L., Imagawa, H., and Tayama, N. (**2004**). "Growl voice in pop and ethnic styles," Proceedings of the International Symposium on Musical Acoustics 2004.
Schwarz, R. (**2007**). "Model-based quantification of pathological voice production," Ph.D. thesis, Universität Erlangen-Nürnberg, Erlangen, Germany.
Stevens, K. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**(3), 82–91.
Švec, J. G., and Schutte, H. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J. Voice **10**, 201–205.
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–3749.
Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.
Traill, A. (**1986**). "The laryngeal sphincter as a phonatory mechanism in !Xóõ (Bushman)," in *Variation, Culture and Evolution in African Populations: Papers in Honour of Dr. Hertha de Villiers*, edited by R. Singer and J. K. Lundy (Witwatersrand University Press, Johannesburg), pp. 123–131.

How stable are acoustic metrics of contrastive speech rhythm?

Lukas Wiget, Laurence White,[a] Barbara Schuppler, Izabelle Grenon, Olesya Rauch, and Sven L. Mattys
Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom

(Received 10 December 2008; revised 22 December 2009; accepted 22 December 2009)

Acoustic metrics of contrastive speech rhythm, based on vocalic and intervocalic interval durations, are intended to capture stable typological differences between languages. They should consequently be robust to variation between speakers, sentence materials, and measurers. This paper assesses the impact of these sources of variation on the metrics %V (proportion of utterance comprised of vocalic intervals), VarcoV (rate-normalized standard deviation of vocalic interval duration), and nPVI-V (a measure of the durational variability between successive pairs of vocalic intervals). Five measurers analyzed the same corpus of speech: five sentences read by six speakers of Standard Southern British English. Differences between sentences were responsible for the greatest variation in rhythm scores. Inter-speaker differences were also a source of significant variability. However, there was relatively little variation due to segmentation differences between measurers following an agreed protocol. An automated phone alignment process was also used: Rhythm scores thus derived showed good agreement with the human measurers. A number of recommendations for researchers wishing to exploit contrastive rhythm metrics are offered in conclusion.
 [DOI: 10.1121/1.3293004]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Kv, 43.70.Mn [MSS] Pages: 1559–1569

I. INTRODUCTION

A. Metrics of contrastive speech rhythm

The study of speech rhythm, once a search for isochronous units of speech (e.g., Lehiste, 1977), has more recently focused on cross-linguistic variation in durational contrast between stressed and unstressed syllables. Dauer (1983) observed that certain phonetic and phonotactic regularities of syllable construction tend to co-occur between languages, at least when one considers the Romance and Germanic languages of Western Europe. Thus, for example, Spanish, French, and Italian have relatively limited clustering of consonants in onsets and codas, with open consonant-vowel syllables being the predominant pattern, whereas English, Dutch, and German allow more complex consonant clusters, with stress tending to occur on heavy syllables, such as those with complex codas. Furthermore, although all of these Western European languages show lengthening of vowels in stressed syllables, Romance languages—Spanish being the most clear-cut example—have attenuated stress-related vowel lengthening compared to the Germanic languages.

Various acoustic metrics have been devised based on these observations (see Table I for definitions of the metrics discussed here), all relying on measurement of the duration of vocalic and consonantal intervals within utterances (e.g., Ramus *et al.*, 1999; Low *et al.*, 2000). Being acoustically based, these intervals take no account of phonological structure, so that coda consonants and immediately following onset consonants are included in the same interval, as are adjacent heterosyllabic vowels (see Ramus *et al.*, 1999, for a discussion of the rationale for this procedure).

These rhythm metrics are intended to capture the stable differences between and within languages in degree of temporal stress contrast, as elucidated by Dauer (1983). Given this assumption, such metrics are best at gauging "contrastive" rhythm, i.e., the balance of strong and weak elements in speech, rather than "dynamic" speech timing, i.e., the temporal arrangement of groups of sounds according to a higher-level structure.

White and Mattys (2007a, 2007b) compared the efficacy of various contrastive rhythm metrics in discriminating between languages held to differ in degree of temporal stress contrast. They also looked for evidence of such differences between varieties of English. In those studies, the most effective metrics were found to be %V (the proportion of total utterance duration made up of vocalic rather than consonantal intervals) and VarcoV (the coefficient of variation of vocalic interval duration, i.e., the standard deviation divided by mean vocalic interval duration, to normalize for articulation rate). Another rate-normalized metric of vocalic interval duration, nPVI-V (Low *et al.*, 2000—see Table I), was also useful and, unsurprisingly, highly correlated with VarcoV, though it manifested somewhat less discriminatory power.

To illustrate the interpretation of contrastive rhythm metrics, Fig. 1 summarizes results for VarcoV and %V from these and related studies (White and Mattys, 2007a, 2007b; White *et al.*, 2009). It is assumed that the high VarcoV scores for English, particularly standard Southern British English (SSBE), and for standard Dutch are a reflection of the strong marking of stress by vowel lengthening in these languages. The lower VarcoV scores for French, Italian, and—in particular—Spanish are assumed to reflect smaller durational

[a]Author to whom correspondence should be addressed. Electronic mail: laurence.white@bristol.ac.uk

TABLE I. The rhythm metrics considered in the present study.

Metric	Description	Main references
ΔV	Standard deviation of vocalic interval duration.	Ramus *et al.* (1999)
ΔC	Standard deviation of consonantal interval duration.	Ramus *et al.* (1999)
%V	Percent of total utterance duration composed of vocalic intervals.	Ramus *et al.* (1999)
VarcoV	Coefficient of variation of vocalic interval duration (i.e., standard deviation of vocalic interval duration divided by the mean), multiplied by 100.	Dellwo (2006); White and Mattys (2007a, 2007b)
VarcoC	Coefficient of variation of consonantal interval duration (i.e., standard deviation of consonantal interval duration divided by the mean), multiplied by 100.	Dellwo (2006); White and Mattys (2007a, 2007b)
nPVI-V	Normalized pairwise variability index for vocalic intervals. Mean of the differences between successive vocalic intervals divided by their sum, multiplied by 100.	Low *et al.* (2000); Grabe and Low (2002)
rPVI-C	Pairwise variability index for consonantal intervals. Mean of the differences between successive consonantal intervals.	Low *et al.* (2000); Grabe and Low (2002)
nPVI-VC	Normalized pairwise variability index for summed vocalic and consonantal intervals. Mean of the differences between successive vocalic+consonantal intervals divided by their sum, multiplied by 100.	Liss *et al.* (2009)

differences between stressed and unstressed vowels. English and Dutch score low on %V, which is interpreted as indicating the preponderance of consonant clusters compared with the Romance languages. Naturally, %V must also be influenced by patterns of vowel duration, and the shortness of unstressed vowels in Dutch and English is assumed also to contribute to their low %V scores. The influence of vowel duration on both metrics is one factor underpinning the negative correlation between %V and VarcoV, together with the fact that—at least in the languages studied thus far—the trends for vowels and consonants tend to co-occur, as described by Dauer (1983). Thus, the overall picture from such studies is of a gradient variation in temporal stress contrast from low (Castilian Spanish) to high (standard Southern British English).

B. Comparison of contrastive rhythm scores between studies

In recent years, contrastive rhythm metrics have been increasingly applied to address a range of issues: E.g., Carter (2005) looked at the influence of Spanish on the rhythm of Hispanic American English; Patel *et al.* (2006) tested the notion that musical rhythm reflects the rhythm of a culture's native language; and Liss *et al.* (2009) attempted to discriminate between different types of dysarthric speech. One significant problem in interpreting and comparing such results is that absolute rhythm scores can vary widely for the same languages between published studies. For example, in the studies of Ramus (2002), Grabe and Low (2002), and White and Mattys (2007a), nPVI-V ranged from 30 to 42 for Castilian Spanish and from 63 to 73 for standard Southern British English; for rPVI-C, Grabe and Low's (2002) score for Spanish was comparable (58) to Ramus' (2002) score for English (57), despite the strong assumption of much greater consonantal interval variation in English (see Table I for definitions of these metrics).

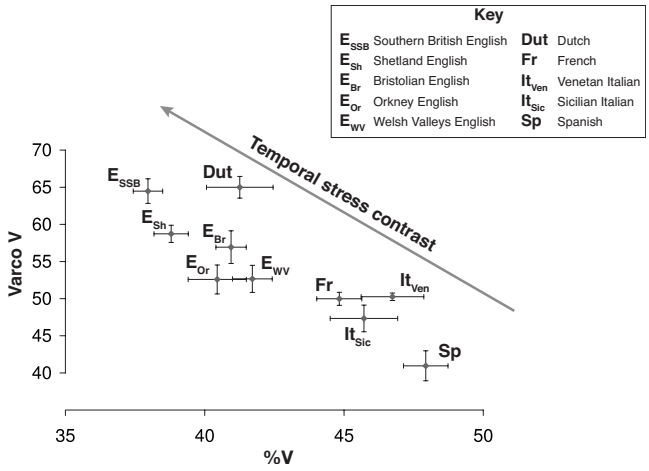


FIG. 1. VarcoV by %V plot showing the scores for Dutch (White and Mattys, 2007a), French (White and Mattys, 2007a), Spanish (White and Mattys, 2007a), two varieties of Italian (White *et al.*, 2009), and several varieties of English (White and Mattys, 2007b).

If contrastive rhythm metrics, as stated above, are intended to capture stable differences between and within languages, such variation in scores represents a significant empirical and theoretical problem. This paper is intended to ascertain whether the variation is merely procedural and can be controlled through improved experimental design, or whether it represents a challenge either to the assumption that languages have stable contrastive rhythmic properties and/or to the assumption that the commonly used rhythm metrics reflect these contrastive rhythmic properties.

There are several obvious potential sources of variation in rhythm scores between studies: The speakers measured, the linguistic materials spoken, and the protocols used for measuring interval duration, together with individual differences between measurers in the application of these protocols. Here we assess the relative variation due to each potential source.

With regard to speakers, there will clearly be variation within any linguistic community in the realization of stress contrasts, as with any segmental or suprasegmental feature, but these differences should be relatively small in comparison with those between languages held to have differing degrees of temporal stress contrast. Variation in articulation

rate is one important factor that might cause significant deviations even within a fairly homogenous accent group, and it is known that certain non-normalized metrics of variation in vowel and consonant interval duration—ΔV, ΔC, rPVI-C—manifest inverse relationships with articulation rate (i.e., higher scores at lower rates—White and Mattys, 2007a). The metrics with most discriminative power—VarcoV, %V, and nPVI-V as discussed above—are robust to variation in rate, however, and so this potential source of between-speaker variation should not be an issue.

Studies also differ in the speech materials used. Given that rhythm metrics are, at least in part, a reflection of phonological structure, variation in rhythm scores is to be expected according to the linguistic materials recorded. In the extreme case, it is possible, for example, to construct English sentences without consonant clusters, so that they mimic the predominant consonant-vowel-consonant-vowel alternation of a Romance language such as Spanish. In this artificial situation, scores for metrics of consonantal interval variation are indeed more similar for English and Spanish than with more natural unconstrained sentences, as are %V scores (Prieto *et al.*, 2009). Of course, while such contrived sentences serve to make a particular point about rhythm metrics, the ideal materials for contrastive rhythm studies should be a representative reflection of the phonological structures of the languages under consideration. The usual—implicit—solution to this problem has been to use a pseudo-random set of naturalistic sentences, although studies such as that of White and Mattys (2007a) have eliminated approximants from the sentence materials in the interest of consistent application of criteria for the identification of segment boundaries.

With regard to measurers, the extraction of interval durations from speech relies mostly on manual inspection of waveforms and spectrograms, and is therefore subject to the vagaries of individual measurers, who may use different criteria or apply common criteria idiosyncratically. The use of experienced phonetically trained measurers and agreement on a set of measurement criteria between studies (see, for example, Turk *et al.*, 2006) are clear first steps toward the maintenance of consistency. Avoidance of awkward segmental junctures—e.g., approximant-vowel or vowel-approximant boundaries—may also help. This study aims to determine the degree of variability in rhythm metric scores between experienced measurers working to a common segmentation protocol.

One potentially useful method of eliminating inter-measurer variability in speech timing research is the use of an automated phone alignment process based on speech recognition algorithms. In studies using read speech, the recognition software is provided with an orthographic or broad phonetic transcription of the sentences, and identifies the boundaries assumed to be present in the signal given this transcription. The practical benefits of such an approach are that a very large amount of speech can be analyzed at little time cost and with consistency of application of segmentation criteria, in contrast with the laborious and error-prone manual approach. Here we examine whether current methods of automated phone alignment—based on statistical properties of the signal rather than phonetically defined criteria—are sufficiently reliable to provide consistency of rhythm scores compared to trained human measurers applying agreed phonetic-based protocols.

C. Purpose of experiment

In this paper, we investigate the robustness of various rhythm metrics to variation in speakers, materials, and measurers, based on an analysis of six speakers of SSBE reading five scripted sentences each. The SSBE sentences were segmented into vocalic and intervocalic intervals by five measurers, all phoneticians trained in the identification of speech segment boundaries. We also used an automatic phone alignment process on the SSBE sentences to derive scores for the same rhythm metrics.

As discussed above, two metrics have been found to be particularly useful (White and Mattys, 2007a, 2007b): VarcoV, the coefficient of variation of vowel interval duration, and %V, the proportion of utterance duration that is vocalic. The focus of the present study is therefore on these two metrics, together with the widely used nPVI-V (normalized pairwise variability index of vowel interval duration), which we found previously to be highly correlated with VarcoV, while being somewhat less discriminative. As a matter of record, however, the primary analyses were carried out for all rhythm metrics listed in Table I, and are reported in the Appendix. We include the results for nPVI-VC, a metric not utilized by White and Mattys (2007a, 2007b), but subsequently found useful in the classification of dysarthric speech by Liss *et al.* (2009).

II. METHOD

A. Participants and measurers

Six speakers, three females and three males, of Standard Southern British English were recorded. None reported any speech or hearing problems. The measurers were all trained phoneticians.

B. Materials

The five English sentences were the same as those used by White and Mattys (2007a) and are listed here in Appendix A. The sentences were constructed to avoid the approximants /l/, /r/, /j/, and /w/, for reasons discussed in Sec. I B.

C. Recordings

The sentences were recorded at the University of Bristol as part of a longer session comprising (1) short story reading, (2) map description, and (3) sentence reading (the five sentences in the present study preceded by five others). Participants were given time to rehearse the sentences silently before reading them aloud and were instructed to speak in their normal conversational voice at a rate that felt natural and comfortable. Sentence readings that contained errors or disfluencies were repeated. Recordings, direct to disk at a sampling rate of 32 kHz, were made in a sound-attenuated room using a high-quality microphone.

D. Measurements

The boundaries between vocalic and consonantal intervals of the Standard Southern British English sentences were labeled by five phonetically trained measurers (M1–M5). Prior to labeling, the measurers agreed to follow the criteria adopted by White and Mattys (2007a). (Note that the results for measurer M1 are those originally reported in White and Mattys, 2007a.) As these criteria were the sole common guide for the measurers in this study, they are reproduced verbatim.

"This procedure was carried out with reference to standard criteria (e.g., Peterson and Lehiste, 1960); where labels were associated with the start or end of pitch periods, they were placed at the point of zero crossing on the waveform.

The primary determiner of the placement of a vowel-consonant boundary was the end of the pitch period preceding a break in formant structure associated with a significant drop in waveform amplitude. Additional criteria which facilitated the location of the boundary in certain contexts included:

- Where the vowel offset was glottalized, a change in the shape of successive pitch periods, for example, lengthening or doubling.
- Before fricatives, the onset of visible frication.
- Before nasals, the appearance of nasal formant structure and a waveform amplitude minimum.

The consonant-vowel boundary was the beginning of the pitch period at the onset of vocalic formant structure, where this was associated with the appearance of pitch periods consistent with the body of the vowel (e.g., unfricated and of comparable amplitude). Aspiration following stop release was therefore included within the consonantal interval." (White and Mattys, 2007a, pp. 506–507).

Following the practice in White and Mattys, 2007a, pauses and disfluencies within utterances were excluded and vowel-vowel or consonant-consonant intervals on either side of the pause were combined. The measurers did not discuss specific labeling criteria, nor did they examine each other's measurements at any stage prior to extraction and analysis of the resulting durational data. These precautions were intended to ensure that the labeling differences between measurers can be regarded as representative of the differences that occur when independent studies use comparable labeling criteria, but different measurers.

The recorded utterances were also subjected to an automatic phone alignment process. Based on an orthographic transcription of each sentence, a forced alignment was carried out with the open-source speech recognition tool SPRAAK (Demuynck *et al.*, 2008). Forty-one tri-state monophonic Gaussian mixture models were trained on the British English SpeechDat FDB telephone speech database (Draxler *et al.*, 1998), separately for male and female speakers. As a 32 ms Hamming window with a step size of 10 ms was used, the duration of the aligned segments is in multiples of 10 ms with a minimum duration of 30 ms. The lexicon used in the forced alignment process was derived from the orthographic transcription using the CELEX lexical database (Baayen *et al.*, 1995). The resulting phone alignments were then transformed into alternations of vocalic and intervocalic intervals, in order to bring them into the same format as the segmentations by the human measurers.

Scores for the metrics %V, ΔV, ΔC, VarcoV, VarcoC, nPVI-V, rPVI-C, and nPVI-VC (see Table I) were then computed, separately for the five read sentences produced by each of the six speakers. For the sake of consistency, utterance-initial consonants, where they occurred, were excluded from the measurements.

E. Statistical analysis

We report on the variability in contrastive rhythm scores according to speaker, sentence materials, and measurers. For the analysis of the variability due to speakers, each speaker contributed 30 scores to each rhythm metric (five sentences and six measurers). For the analysis of the variability due to sentence materials, each sentence contributed 36 scores to each rhythm metric (six speakers and six measurers). For the analysis of the variability due to measurers, each measurer produced 30 scores per rhythm metric (six speakers and five sentences). The intraclass correlation coefficients (ICCs) were computed with the R environment's IRR package (Gamer *et al.*, 2007). We used a two-way model, which treats both raters and items as randomly sampled from a population (Shrout and Fleiss, 1979). Intraclass correlation coefficients of 0.40–0.59 are commonly regarded as indicating moderate inter-rater reliability, values of 0.60–0.79 as substantial, and values of 0.80 or larger as outstanding (Garson, 2009). In addition, in order to provide some sense of whether differences between speakers, sentences, and measurers are potentially problematic for interpretation, we compared the size of those differences to the largest difference reported between languages in White and Mattys (2007a) as a benchmark, namely, the difference between scores for Castilian Spanish and Standard Southern British English (see Fig. 1 for VarcoV and %V values).

III. RESULTS AND DISCUSSION

As discussed above, we focus on the rhythm metrics found by previous studies (White and Mattys, 2007a, 2007b) to be the most discriminative between and within languages: VarcoV, the coefficient of variation of vocalic interval duration (standard deviation divided by the mean); nPVI, the normalized pairwise variability index of vocalic interval duration (see Table I for fuller definition); and %V, the proportion of total utterance duration made up of vocalic rather than consonantal intervals.

A. Effect of speaker on %V, VarcoV, and nPVI-V scores

Figure 2 shows the mean %V, VarcoV, and nPVI-V scores of the six Standard Southern British English speakers averaged across sentences and measurers. The means for the other rhythm metrics can be found in Appendix B.

A one-way repeated measures analysis of variance (ANOVA) for %V showed a main effect of Speaker, $F(5,145)=15.53$, $p<0.001$. Values for ICCs and 95% con-

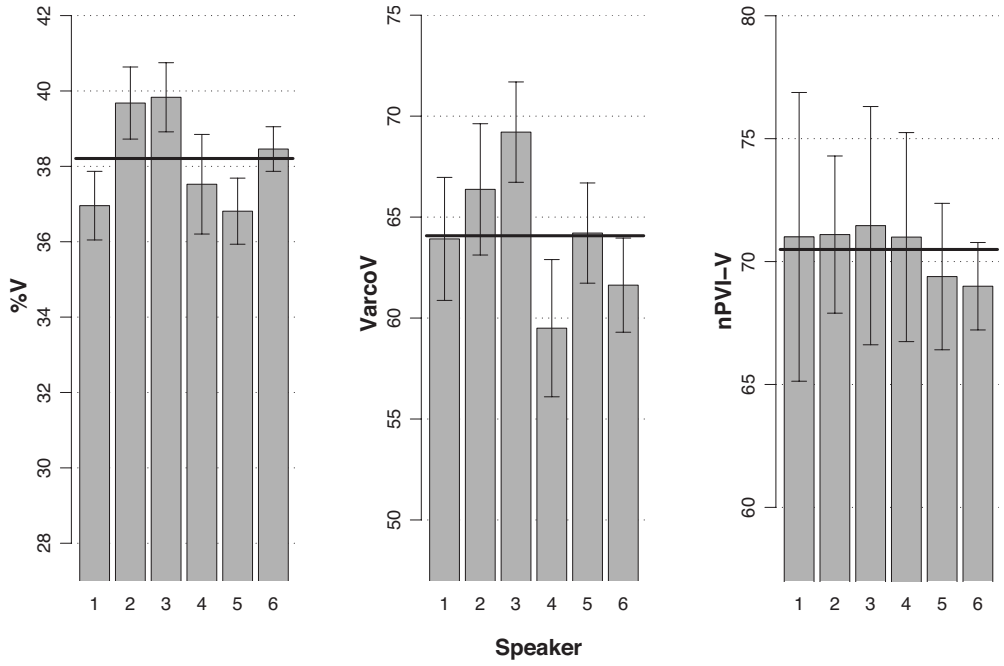


FIG. 2. Means and standard errors of the rhythm metrics %V, VarcoV, and nPVI-V for the six SSBE speakers. The solid horizontal bars indicate the pooled means.

fidence intervals for all metrics and conditions are summarized in Table II. As can be seen, consistency among speakers for %V was moderate. The largest difference in mean %V was 3.0 or 30% of the Spanish-English difference and the mean difference was 1.6 or 16% of Spanish-English. For VarcoV, there was also a main effect of speaker, $F(5,145)=13.47$, $p<0.001$. Consistency among speakers was moderate (Table II). The largest difference in mean VarcoV was 9.7 or 42% of Spanish-English, and the mean difference was 4.2 or 18% of Spanish-English. For nPVI-V, we found no main effect of speaker, $F(5,145)<1$. Consistency among speakers was again moderate (Table II), but here the actual differences between the speakers' mean scores were very low: The largest difference in mean nPVI-V was 2.5 or only 7% of Spanish-English, and the mean difference was 1.2 or 3% of Spanish-English.

TABLE II. Intraclass correlation coefficients, among speakers, sentences, and measurers, for the rhythm metrics %V, VarcoV, and nPVI-V (with 95% confidence intervals). Values of 0.40–0.59 are commonly regarded as indicating moderate inter-rater reliability, values of 0.60–0.79 as substantial, and values of 0.80 or larger as outstanding (Garson, 2009).

	%V	VarcoV	nPVI-V
1. Speakers	0.53 [0.38,0.70]	0.56 [0.41,0.72]	0.52 [0.36,0.68]
2. Sentences	0.38 [0.23,0.55]	0.32 [0.17,0.49]	0.01 [−0.08,0.15]
3a. Measurers (including automatic alignment)	0.64 [0.49,0.77]	0.64 [0.49,0.77]	0.52 [0.36,0.68]
3b. Measurers (excluding automatic alignment)	0.67 [0.52,0.80]	0.71 [0.57,0.82]	0.62 [0.47,0.76]

All other rhythm metrics also showed a main effect of speaker (Appendix B). The ICCs of ΔV and nPVI-VC were in the range of those for %V, VarcoV, and nPVI-V; the ICCs for the consonantal interval metrics were lower.

B. Effect of sentence on %V, VarcoV, and nPVI-V scores

Figure 3 shows the mean %V, VarcoV, and nPVI-V scores of the five English sentences averaged across speakers and measurers. The means for the other rhythm metrics can be found in Appendix C.

For %V, a one-way repeated measures ANOVA showed a main effect of sentence, $F(4,140)=27.10$, $p<0.001$. Consistency between sentences, as measured by the ICC, was low (Table II). The largest difference in mean %V between sentences was 4.9 or 49% of Spanish-English and the mean difference was 2.0 or 20% of Spanish-English. For VarcoV, there was also a main effect of sentence, $F(4,140)=35.38$, $p<0.001$. Consistency between sentences was low (Table II). The largest difference in mean VarcoV was 13.0 or 57% of Spanish-English, and the mean difference was 6.5 or 28% of Spanish-English. For nPVI-V, we also found a main effect of sentence, $F(4,140)=35.49$, $p<0.001$. Consistency between sentences was very low (Table II). The largest difference in mean nPVI-V was 15.5 or 42% of Spanish-English, and the mean difference was 9.0 or 25% of Spanish-English.

All other rhythm metrics also showed a main effect of Sentence (Appendix C). The ICCs likewise all indicated a low consistency between sentences. As with the analysis of speaker differences above, the ICCs of the consonantal interval metrics were generally lower than those of the vocalic metrics, with the exception of a very low value for nPVI-V.

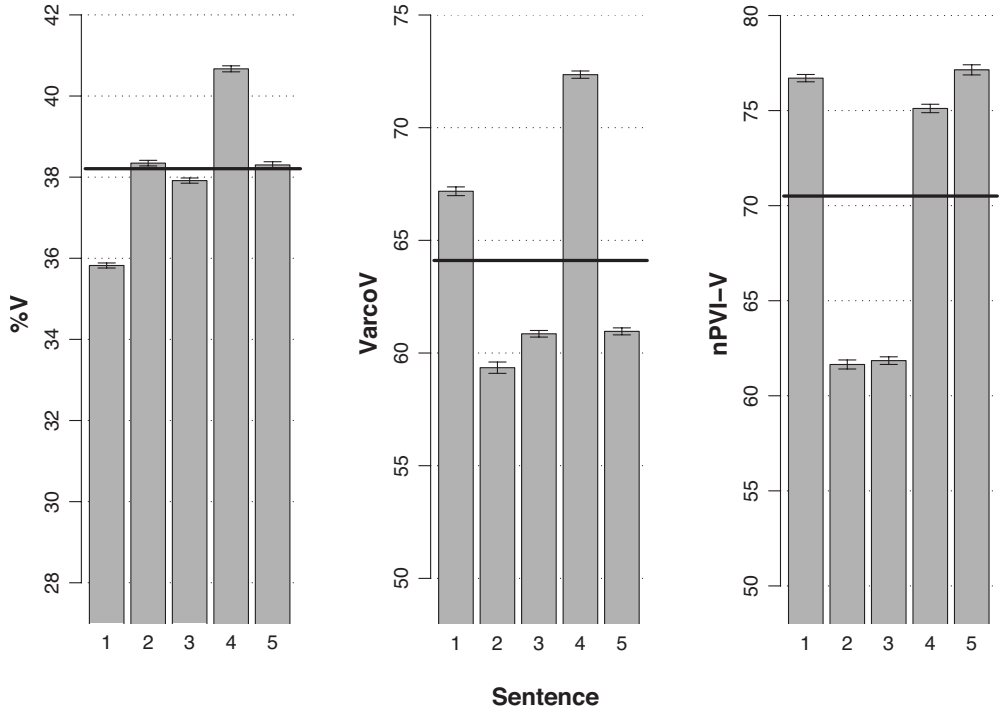


FIG. 3. Means and standard errors of the rhythm metrics %V, VarcoV, and nPVI-V for the five English sentences as read by the SSBE speakers. The solid horizontal bars indicate the pooled means.

C. Effect of measurer on %V, VarcoV, and nPVI-V scores

Figure 4 shows the mean %V, VarcoV, and nPVI-V scores of the five human measurers (M1–M5) and the automatic phone alignment (A) averaged across speakers and sentences. The means for the other rhythm metrics can be found in Appendix D.

For %V, a one-way repeated measures ANOVA showed a main effect of measurer, $F(5,145)=15.85$, $p<0.001$. Consistency among measurers, as measured by the ICC, was, however, substantial (Table II). The largest difference in mean %V was 3.7, or 37% of Spanish-English, and the mean difference was 1.4 or 14% of Spanish-English. For VarcoV, we also found a main effect of measurer, $F(5,145)=10.06$, $p<0.001$. Consistency among measurers was substantial (Table II). The largest difference in mean VarcoV was 6.4 or 28% of Spanish-English, and the mean difference was 3.4 or 15% of Spanish-English. For nPVI-V, there was a main effect of measurer, $F(5,145)=2.46$, $p<0.05$. Consistency among measurers was moderate (Table II). The largest dif-

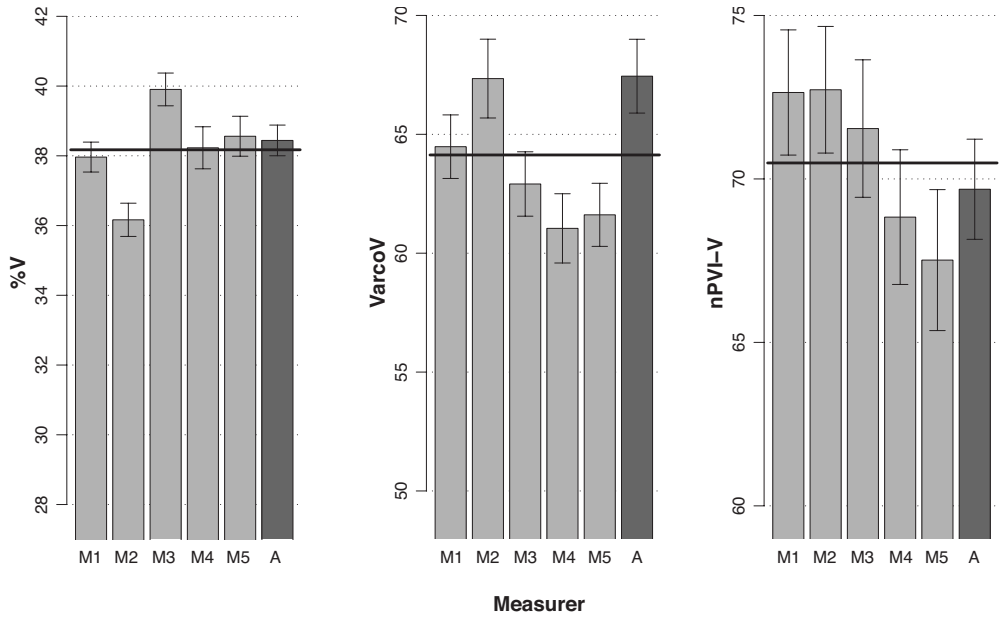


FIG. 4. Means and standard errors of the rhythm metrics %V, VarcoV, and nPVI-V (for SSBE) for the five human measurers (M1–M5) and the automatic phone alignment (A). The solid horizontal bars indicate the pooled means.

ference in mean nPVI-V was 5.2 or 14% of Spanish-English difference, and the mean difference was 2.6 or 7% of Spanish-English.

All other rhythm metrics also showed a main effect of measurer (Appendix D), with the exception of ΔV, where the effect approached significance. The consonantal interval metrics were again generally less consistent across measurers than the vocalic interval metrics.

As for the performance of the automatic phone alignment, Fig. 4 shows that the mean %V and nPVI-V scores were well within the range of the human measurers, and for VarcoV were just outside the human measurers' range. This observation was confirmed by pairwise comparisons between the automatic alignment and mean values for the five human measurers. There was no difference in mean for %V, $t(29)=0.08$, $p=0.44$, and nPVI-V, $t(29)=-0.51$, $p=0.61$. For VarcoV, the mean score of the automatic alignment was higher than that of the human measurers, $t(29)=3.02$, $p=0.005$. When we compared the consistency among human measurers with that among all measurers (i.e., including the automatic alignment) we found that the ICCs among the human measurers alone are somewhat higher (Table II); however, the overlapping confidence intervals indicate that these differences were not significant.

D. Discussion

The analyses show that variation between speakers, sentences, and measurers all contribute to variability in rhythm metric scores. Sentences produced the largest variation in scores, with, for example, the mean difference in VarcoV between sentences being 28% of the Spanish-English difference. The differences due to speakers and measurers were generally smaller. In addition, consistency was highest among measurers (generally with an ICC of above 0.60) and lowest between sentences (ICC below 0.40). Between speakers, the ICC was around 0.50. The confidence intervals for sentence ICC and measurer ICC did not overlap for either %V or VarcoV, indicating that, for these two metrics, consistency among sentences was significantly lower than consistency among measurers. With regard to individual rhythm metrics, differences were generally smallest for nPVI-V and largest for VarcoV. Consistency, on the other hand, was slightly lower for nPVI-V than for the other two metrics: This likely to be in part the result of the lack of variation in nPVI-V between speakers; with little variation in the other two variables, there is little scope for sentences to show consistency in their patterns of scores.

How relevant are these differences to the comparison of results between studies? The largest difference between the English sentences examined here, for example, was more than half the size (57%) of the mean English-Spanish difference, which can be considered the archetypal poles of the contrastive rhythm space, at least among languages studied so far. Thus, a small unrepresentative set of sentences could easily alter the position of the language measured within the rhythm space, as also demonstrated by Prieto *et al.* (2009).

What then are "representative" sentences for the purpose of rhythm analysis? At the segmental level, Prieto *et al.* (2009) clearly showed that the phonological structure of syllables has an impact on rhythm scores, with untypical patterns (such as sentences comprised of a succession of open syllables for English) producing misleading results. At the suprasegmental level, one critical factor is clearly the relative numbers and distribution of strong and weak syllables. Low *et al.* (2000) constructed sentences containing only strong syllables and showed that nPVI-V scores were much lower than those for sentences containing an alternating pattern of strong and weak syllables.

TABLE III. The number of syllable transitions without a change in stress (weak-weak or strong-strong), the number of total transitions, and the CRI for the English sentences.

Sentence	Constant transitions	Total transitions	Constant regularity index
1	4	14	0.71
2	6	16	0.63
3	5	14	0.64
4	4	16	0.75
5	2	16	0.88

As a further test of the effect of suprasegmental structure on rhythm measures, we introduce the contrast regularity index (CRI), a measure of the regularity of sentence stress patterns. We calculated CRI as $1-(c/t)$, where c stands for the number of syllable transitions in which the stress remains constant (a weak syllable followed by a weak syllable or a strong syllable followed by another strong syllable), and t for the total number of syllable transitions. The lower the number of syllable-to-syllable transitions where stress is unaltered, the higher the CRI, with a CRI of 1 indicating a strict alternation of strong and weak syllables.

The CRIs for the five English sentences used in this study are listed in Table III. Given that the PVI was devised to reflect the strength of the alternation between successive syllables, it is reassuring that patterns of CRI values quite closely correspond to the nPVI-V values, with the two sentences lowest in CRI (sentences 2 and 3) also having the lowest nPVI-V scores, and the sentences with a high CRI (sentences 1, 4, and 5) having the highest nPVI-V scores.

It is worth noting that, although VarcoV scores were generally highly correlated with nPVI scores, the pattern of scores was very different for sentence 5, which had a low VarcoV score and yet the highest nPVI-V score. This was also the sentence that had the highest CRI value, reflecting its regularly alternating strong-weak pattern, only disrupted by two weak-weak transitions. The ordering of strong and weak syllables is thus critical for nPVI-V, and irrelevant for simple global measures of interval duration variation such as VarcoV, as discussed in Low *et al.*, 2000. This sensitivity to the rhythmic structure of utterances could be seen as a strength of the PVI, but could also be a weakness in studies that lack a sufficiently large and language-typical range of metrical structure.

In summary, it is clear that all three factors—speakers, sentences, and measurers—need to be considered when developing studies and comparing results. We found that the largest source of variation in rhythm scores was the sentences selected for the study: These can either be randomly

sampled or specifically designed to be representative of the language in terms of phonological and suprasegmental structures, as discussed above. Speakers—a moderate source of variation in rhythm scores—can be naturally treated as a random factor and so the usual caveat regarding low speaker numbers applies. There is little corrective action available, however, regarding idiosyncratic differences between measurers apart from the obvious step of ensuring that all measurers conform to an agreed protocol for the identification of segment boundaries. Given that, it is reassuring to note that differences in rhythm scores between measurers are relatively small compared to those arising from speakers and, in particular, sentences, and that consistency between measurers—as indexed by intraclass correlation coefficients—is relatively high.

IV. GENERAL DISCUSSION

In an attempt to assess the reliability of contrastive rhythm metrics used routinely in phonetic research, we examined how such metrics are affected by variation between measurers, sentences, and speakers. We analyzed five English sentences spoken by six speakers of Standard Southern British English, with segmentation into vocalic and intervocalic intervals carried out by five phoneticians, together with an automated alignment using speech recognition software.

The results clearly demonstrate that the nature of the sentence materials recorded is the most important source of variation in rhythm scores. Consistency of scores between sentences was low for all metrics, even those metrics—%V, VarcoV, and nPVI-V—previously found to be most discriminative between language groups and robust to variation in articulation rate (White and Mattys, 2007a, 2007b). Mean differences between SSBE sentences for these three metrics were equivalent to 20%–28% of the differences between Castilian Spanish and Standard Southern British English, which were the largest cross-linguistic differences found in our previous study (White and Mattys, 2007a), with %V the most consistent and nPVI-V the least consistent between sentences. Such differences raise potentially serious problems of interpretation of rhythm scores given the underlying assumption that they should reflect stable rhythmical properties of languages. One solution to this problem is to record a sufficiently large number of randomly sampled sentences to provide a representative selection. Alternatively, materials may be constructed to accurately represent the phonological and metrical structures encountered in the natural speech of the languages under investigation. We have proposed a contrast regularity index as a simple statistical measure of metrical structure; other measures would be required to assess phonological patterns such as the clustering of consonants in onsets and codas. Analyzing recordings of natural spontaneous speech represents a third approach to the problem.

Variation in rhythm scores between speakers was not as marked as between sentences, with mean differences for the three metrics less than 18% of the Spanish-English difference, and particularly low for nPVI-V. Consistency between speakers was nevertheless only moderate and provides a strong indication against taking single-speaker studies to be in any way representative of population means. Of course, rhythm metrics may be efficacious for comparing scores within a single speaker under different speaking conditions or assessing changes over time, for example, regarding the deterioration in a speaker's prosody due to speech pathology (e.g., Liss *et al.*, 2009), the efficacy of ameliorative speech therapy, or progress to proficiency in second language acquisition. Assuming that the caveats regarding materials are heeded, the use of rhythm metrics in longitudinal studies such as these is one of the potentially most productive areas of their application.

Inter-measurer reliability was relatively good, with mean differences for %V, VarcoV, and nPVI-V between 7% and 15% of the Spanish-English difference. As with speakers, nPVI-V showed the smallest mean difference between measurers, although intraclass correlation coefficients showed that consistency was numerically somewhat lower than for %V and VarcoV.

A practical implication of these findings is that it should be possible to use several measurers within one study without substantial loss of reliability. Our measurers provided segmentations based on their own interpretation of a set of written instructions without consulting each other. By ensuring that measurers discuss problem cases and update their joint segmentation protocol accordingly, we expect that differences between measurers could be reducible to a nonsignificant level. In addition, of particular practical interest is our finding that the automatic phone alignment produced rhythm metric scores that were comparable to those of the human measurers. Of the three metrics examined here, only for VarcoV was the score of the automatic alignment different from the mean score of the human measurers, although the automatic alignment still performed within the range of the human measurers. Automatic phone alignment could, therefore, be used in future studies of contrastive speech rhythm, assuming an adequate supply of training data for the languages under investigation.

Finally, as this paper is essentially a methodological study, we present some recommendations for researchers intending to utilize contrastive rhythm metrics. The first two recapitulate what we have learnt from our previous work (White and Mattys, 2007a, 2007b); the others summarize the findings of the present study.

(1) %V, VarcoV, and nPVI-V are robust to variation in articulation rate and are effective at discriminating between language varieties previously held to differ in terms of contrastive rhythm. However, as all rhythm metrics have limitations, it is safest to use %V in combination with either VarcoV or nPVI-V rather than rely on a single metric.
(2) Results for non-rate-normalized metrics (ΔV, ΔC, and rPVI-C) are difficult to interpret and not reliably discriminative, likewise all metrics of consonantal interval variation (VarcoC, in addition to the non-rate-normalized ΔC and rPVI-C). Furthermore, they show relatively poor consistency between speakers, sentences, and measurers.
(3) Single-speaker or low-N studies should absolutely be avoided where speakers are intended to be representative

of a particular linguistic group. Contrastive rhythm metrics may be useful in single-speaker longitudinal studies, however.

(4) Rhythm scores are strongly affected by the particular linguistic materials used. Either a large sample of sentences should be used or materials should be constructed to be representative of the relevant phonological and metrical properties of the language under study.
(5) Where several measurers are used, they should work according to an agreed protocol for the identification of segment boundaries. Furthermore, discussion and comparison of difficult cases between measurers, which were avoided in the current study, should help minimize variation in rhythm scores.
(6) Contrastive rhythm scores obtained through automatic alignment show good agreement with those obtained from human measurers, assuming sufficient training data are available for the language in question. Of course, use of automated methods would allow much larger sampling of speakers and sentences.
(7) Do not rely too heavily on contrastive rhythm metrics or over-interpret the results of studies that use them. They merely provide an approximate indication of the degree of temporal stress contrast in a language and are susceptible to extraneous variation from multiple sources.

ACKNOWLEDGMENTS

This research was supported by a grant from the Leverhulme Trust (United Kingdom) to S. L. Mattys (Grant No. F/00182/BG) and a Research Training Network grant from the Marie Curie Foundation (Grant No. MRTN-CT-2006-035561). We thank Esther de Leeuw for help with segmentation.

APPENDIX A: THE SENTENCES READ BY THE SSBE SPEAKERS IN THIS STUDY

The following are the sentences read by the SSBE speakers.

(1) The supermarket chain shut down because of poor management.
(2) Much more money must be donated to make this department succeed.
(3) In this famous coffee shop they serve the best doughnuts in town.
(4) The chairman decided to pave over the shopping center garden.
(5) The standards committee met this afternoon in an open meeting.

APPENDIX B: SPEAKER COMPARISONS

Means (and standard errors) of the rhythm metrics ΔV, ΔC, VarcoC, rPVI-C, and nPVI-VC according to speaker; the ICC among speakers (with 95% confidence intervals); and the *F*-test for a main effect of speaker.

	Sp 1	Sp 2	Sp 3	Sp 4	Sp 5	Sp 6	ICC	*F*-test
ΔV	53.8 (1.4)	55.7 (1.3)	55.5 (0.6)	45.0 (1.3)	42.9 (0.6)	46.4 (1.1)	0.45 [0.29,0.62]	$F(5,145)=48.06$, $p<0.001$
ΔC	69.3 (2.5)	55.1 (2.0)	58.8 (2.4)	60.4 (2.1)	54.0 (2.2)	57.7 (1.4)	0.26 [0.12,0.45]	$F(5,145)=8.94$, $p<0.001$
VarcoC	48.2 (1.6)	42.8 (1.2)	47.8 (1.4)	47.6 (1.1)	46.4 (1.2)	48.2 (1.5)	0.25 [0.12,0.44]	$F(5,145)=3.21$, $p=0.009$
rPVI-C	82.9 (3.8)	66.5 (2.5)	65.7 (3.0)	72.6 (2.8)	66.4 (2.7)	67.9 (2.5)	0.37 [0.22,0.56]	$F(5,145)=8.32$, $p<0.001$
nPVI-VC	40.9 (1.3)	37.2 (1.3)	41.5 (1.7)	40.2 (1.1)	40.9 (1.4)	34.2 (0.8)	0.61 [0.46,0.75]	$F(5,145)=12.32$, $p<0.001$

APPENDIX C: SENTENCE COMPARISONS

Means (and standard errors) of the rhythm metrics ΔV, ΔC, VarcoC, rPVI-C, and nPVI-VC according to sentence read; the ICC among sentences (with 95% confidence intervals); and the *F*-test for a main effect of sentence.

	Sen 1	Sen 2	Sen 3	Sen 4	Sen 5	ICC	*F*-test
ΔV	51.9 (1.3)	48.3 (1.4)	49.8 (1.0)	55.3 (1.3)	44.1 (1.0)	0.49 [0.33,0.65]	$F(4,140)=23.50$, $p<0.001$

ΔC	59.3 (0.9)	65.8 (2.3)	59.4 (1.6)	55.8 (2.7)	55.8 (2.1)	0.22 [0.09,0.40]	$F(4,140)=5.26$, $p<0.001$
VarcoC	43.0 (0.7)	49.9 (1.1)	44.2 (1.0)	49.4 (1.6)	47.6 (1.3)	0.05 [−0.05,0.20]	$F(4,140)=7.07$, $p<0.001$
rPVI-C	63.3 (1.5)	74.6 (3.0)	84.1 (1.8)	65.5 (3.4)	64.1 (2.2)	0.25 [0.11,0.43]	$F(4,140)=16.72$, $p<0.001$
nPVI-VC	41.9 (0.9)	33.9 (0.7)	35.2 (0.8)	36.9 (0.9)	47.8 (1.3)	0.16 [0.04,0.33]	$F(4,140)=44.48$, $p<0.001$

APPENDIX D: MEASURER COMPARISONS

Means (and standard errors) of the rhythm metrics ΔV, ΔC, VarcoC, rPVI-C, and nPVI-VC according to measurer (human measurers M1–M5 and automatic phone alignment A); the ICC among measurers (with 95% confidence intervals); and the *F*-test for a main effect of measurer.

	M1	M2	M3	M4	M5	A	ICC	*F*-test
ΔV	49.2 (1.4)	49.5 (1.4)	50.2 (1.4)	50.1 (1.4)	48.8 (1.5)	51.5 (1.7)	0.79 [0.69,0.88]	$F(5,145)=2.02$, $p=0.080$
ΔC	58.6 (1.9)	59.1 (1.9)	57.0 (1.9)	68.1 (3.1)	58.6 (2.0)	53.9 (1.9)	0.42 [0.27,0.60]	$F(5,145)=8.20$, $p<0.001$
VarcoC	46.8 (1.2)	45.4 (1.2)	47.4 (1.2)	50.8 (1.8)	46.4 (1.2)	44.1 (1.3)	0.43 [0.27,0.61]	$F(5,145)=4.92$, $p<0.001$
rPVI-C	69.9 (2.6)	69.1 (2.5)	68.0 (2.8)	83.1 (4.0)	68.5 (2.8)	63.4 (2.7)	0.54 [0.39,0.70]	$F(5,145)=11.23$, $p<0.001$
nPVI-VC	39.7 (1.5)	38.3 (1.5)	39.6 (1.3)	40.8 (1.5)	39.3 (1.4)	36.9 (0.9)	0.68 [0.55,0.81]	$F(5,145)=3.07$, $p=0.011$

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (**1995**). The CELEX lexical database. Release 2 (CD-ROM) (Linguistic Data Consortium, University of Pennsylvania, PA).

Carter, P. M. (**2005**). "Quantifying rhythmic differences between Spanish, English, and Hispanic English," in *Theoretical and Experimental Approaches to Romance Linguistics: Selected Papers From the 34th Linguistic Symposium on Romance Languages*, Current Issues in Linguistic Theory Vol. **272**, edited by R. S. Gess and E. J. Rubin (John Benjamins, Amsterdam), pp. 63–75.

Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phonetics **11**, 51–62.

Dellwo, V. (**2006**). "Rhythm and speech rate: A variation coefficient for deltaC," in Language and Language Processing: Proceedings of the 38th Linguistic Colloquium, Peter Lang, Frankfurt, pp. 231–241.

Demuynck, K., Roelens, J., Van Compernolle, D., and Wambacq, P. (**2008**). "Spraak: An open source speech recognition and automatic annotation kit," in Interspeech 2008, Brisbane (CDROM).

Draxler, C., van den Heuvel, H., and Tropf, H. (**1998**). "SpeechDat experiences in creating large multilingual speech databases for teleservices," in Proceedings of the LREC98, Granada, pp. 361–366.

Gamer, M., Lemon, J., and Fellows, I. (**2007**). "irr: Various coefficients of interrater reliability and agreement," R package version 0.70.

Garson, G. D. (**2009**). "Reliability analysis," http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm (Last viewed 7/12/2009).

Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.

Lehiste, I. (**1977**). "Isochrony reconsidered," J. Phonetics **5**, 253–256.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S.M., and Caviness, J. N. (**2009**). "Quantifying speech rhythm abnormalities in the dysarthrias," J. Speech Lang. Hear. Res. **52**, 1334–1352.

Low, E. L., Grabe, E., and Nolan, F. (**2000**). "Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English," Lang Speech **43**, 377–401.

Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (**2006**). "Comparing the rhythm and melody of speech and music: The case of British English and French," J. Acoust. Soc. Am. **119**, 3034–3047.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Prieto, P., Vanrell, M. M., Astruc, L., Payne, E., and Post, B. (**2009**). "Top-down planning of rhythm: Evidence from Catalan, English, and Spanish," paper presented at Phonetics and Phonology in Iberia 2009, Las Palmas de Gran Canaria, June.

Ramus, F. (**2002**). "Acoustic correlates of linguistic rhythm: Perspectives," in Proceedings of the Speech Prosody 2002, Aix-en-Provence, pp. 115–120.

Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292.

Shrout, P. E., and Fleiss, J. L. (**1979**). "Intraclass correlations: Uses in assessing rater reliability," Psychol. Bull. **86**, 420–428.

Turk, A., Nakai, S., and Sugahara, M. (**2006**). "Acoustic segment durations in prosodic research: A practical guide," in *Methods in Empirical Prosody Research*, edited by S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schliesser (de Gruyter, Berlin), pp. 1–28.

White, L., and Mattys, S. L. (**2007a**). "Calibrating rhythm: First language and second language studies," J. Phonetics **35**, 501–522.
White, L., and Mattys, S. L. (**2007b**). "Rhythmic typology and variation in first and second languages," in *Segmental and Prosodic Issues in Romance Phonology*, edited by P. Prieto, J. Mascaró, and M.-J. Solé (John Benjamins, Amsterdam), pp. 237–257.
White, L., Payne, E., and Mattys, S. L. (**2009**). "Rhythmic and prosodic contrast in Venetan and Sicilian Italian," in *Phonetics and Phonology: Interactions and Interrelations*, edited by M. Vigario, S. Frota, and M. J. Freitas (John Benjamins, Amsterdam), pp. 137–158.

Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing

Koenraad S. Rhebergen[a)]
Department of Clinical and Experimental Audiology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Johannes Lyzenga
VU University Medical Center, ENT/Audiology, EMGO Institute, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

Wouter A. Dreschler
Department of Clinical and Experimental Audiology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Joost M. Festen
VU University Medical Center, ENT/Audiology, EMGO Institute, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 23 April 2008; revised 15 December 2009; accepted 17 December 2009)

The speech intelligibility index (SII) is an often used calculation method for estimating the proportion of audible speech in noise. For speech reception thresholds (SRTs), measured in normally hearing listeners using various types of stationary noise, this model predicts a fairly constant speech proportion of about 0.33, necessary for Dutch sentence intelligibility. However, when the SII model is applied for SRTs in quiet, the estimated speech proportions are often higher, and show a larger inter-subject variability, than found for speech in noise near normal speech levels [65 dB sound pressure level (SPL)]. The present model attempts to alleviate this problem by including cochlear compression. It is based on a loudness model for normally hearing and hearing-impaired listeners of Moore and Glasberg [(2004). Hear. Res. **188**, 70–88]. It estimates internal excitation levels for speech and noise and then calculates the proportion of speech above noise and threshold using similar spectral weighting as used in the SII. The present model and the standard SII were used to predict SII values in quiet and in stationary noise for normally hearing and hearing-impaired listeners. The present model predicted SIIs for three listener types (normal hearing, noise-induced, and age-induced hearing loss) with markedly less variability than the standard SII.
 [DOI: 10.1121/1.3291000]

PACS number(s): 43.71.An, 43.66.Ba, 43.71.Es, 43.71.Ky [KWG] Pages: 1570–1583

I. INTRODUCTION

One of the first studies on modeling speech intelligibility was started at AT&T Bell Laboratories in the 1920s. The aim of these studies was to analyze speech quality, telephone quality, and speech transmission. The work was published internationally only shortly after World War II (French and Steinberg, 1947; Fletcher and Galt, 1950; Fletcher, 1953), and formed the basis of the "articulation theory" (Allen, 1996). The principle of the articulation scheme was to calculate the available speech information at the receiver's ear (listener), and the outcome was a number between zero and unity, called the articulation index (AI). An AI of zero meant that no information was present at the receiver's ear, whereas an AI of unity meant that all the information was transmitted properly. Unfortunately, the AI method was hardly used after its publication (Müsch, 2001), which may well have been due to its complex calculations. To overcome this problem, Kryter (1962) introduced a calculation scheme that was based on the articulation theory, but much simpler to use. Eventually, his method led to the standard ANSI S3.5 (1969) AI calculation method. The AI model kept on evolving, and today a revised version, named the speech intelligibility index (SII) (ANSI S3.5-1997, 1997), is commonly used (e.g., Noordhoek *et al.*, 2000; Hornsby and Ricketts, 2003; Kates and Arehart, 2005; Dubno *et al.*, 2002; Woods *et al.*, 2006). The SII model was designed to predict speech intelligibility for conditions with noise and through communication channels. Nevertheless, it is often used for hearing aid settings and adjustments, and for the evaluation of hearing aid benefit.

The basic principle of the SII model is as follows: For a given condition where speech is embedded in noise at a given signal-to-noise ratio (SNR), the SII is calculated using the average noise spectrum, the average speech spectrum, and the listener's hearing threshold. By calculating the SNR in different frequency bands, and by assigning a weight to each band (the so-called band-importance function, e.g., Pavlovic, 1994, the SII is finally composed by adding the frac-

a)Author to whom correspondence should be addressed. Electronic mail: k.s.rhebergen@amc.uva.nl

tions of available speech in each band (e.g., Studebaker *et al.*, 1993), while accounting for distorted speech at higher levels [>73 dB sound pressure level (SPL)] with aid of a level distortion factor (e.g., Hou and Thornton, 1994; Sherbecoe and Studebaker, 2003). As with the AI, the SII is a number between zero and unity and can been interpreted as the proportion of audible speech that is available to the listener. The articulation theory assumes a fixed relation between the proportion of audible speech and the observed performance for normal-hearing (NH) listeners in their native language, regardless of the type of listening condition (i.e., in quiet or in background noise). The shape of the appropriate transfer function (e.g., Sherbecoe and Studebaker, 1990, 2003, 2002) depends on the speech material (e.g., words and sentences) and the measurement method used (word score, sentence score, up-down procedure, and fixed levels). Such transfer functions can be used in corresponding listening conditions to predict speech intelligibility as a function of speech level in quiet or in noise, or conversely, to describe the observed performance [e.g., the speech reception threshold (SRT); Plomp and Mimpen, 1979] for a chosen proportion of audible speech (i.e., a chosen SII). For example, the proportion of audible speech required to understand about half of a list of short, everyday, Dutch sentences presented in stationary noise is about one-third (SII=0.33). This is true in native normally hearing listeners for their SRT for Dutch sentences (Noordhoek *et al.*, 2000, 2001; van Schijndel *et al.*, 2001; Versfeld and Dreschler, 2002; Rhebergen and Versfeld, 2005). In addition, when using the extended SII (ESII) model (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006) the same proportion is predicted for sentences in fluctuating noise. The extended SII predicts speech intelligibility for sentences in fluctuating (real-life) noise with reasonably good accuracy (Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2006, 2008; George *et al.*, 2006, 2007). The relationship between the SII value and the SRT for hearing-impaired (HI) listeners is less clear. Some hearing-impaired listeners have SII values of about one-third at the observed SRTs whereas others have marked higher SII values at the SRT. Apparently, this latter group of hearing-impaired listeners require more audible speech (i.e., SII >0.33) at the SRT than do normally hearing listeners. An elevated SII associated with an observed SRT is often associated with a supra-threshold deficit in such listeners (e.g., Noordhoek *et al.*, 2000, 2001; van Schijndel *et al.*, 2001; Versfeld and Dreschler, 2002; Festen and Plomp, 1990; George *et al.*, 2006). For such listeners the SII model is not able to explain the observed SRTs on the basis of audibility alone. Factors other than raised hearing levels likely contribute to the observed SRTs (e.g., decreased temporal and spectral resolution).

Unfortunately, the SII model is less successful in predicting the SRT in quiet (e.g., Holube and Kollmeier, 1996; Peters *et al.*, 1998; Noordhoek *et al.*, 2000, 2001). For normally hearing listeners, SII values for the SRT in quiet (SRTq) are often found to be higher than for the SRT in stationary noise (SRTn) (e.g., Noordhoek *et al.*, 2000, 2001). As the performance (number of sentences repeated correctly) is the same for the SRTq and the SRTn, this implies that the SII model predicts that a larger proportion of audible speech (i.e., a higher SII) is required to understand half of the sentences in quiet relative to in noise. Apart from possible effects of phonemic restoration in noise, this is rather unlikely because the absolute threshold is often considered to be an internal noise floor (French and Steinberg, 1947; ANSI S3.5-1997, 1997; ANSI S3.4-2007, 2007). Furthermore, the variance across listeners of the calculated SIIs for SRTq data is often larger than found for SRTn data (e.g., Noordhoek *et al.*, 2001), which, on the other hand, may also be partly attributed to measurement errors and variability of the individual hearing thresholds. Lastly, the SII differences between SRTq data for normally hearing and hearing-impaired listeners were larger than found between those groups for SRTn data (e.g., Noordhoek *et al.*, 2001). A preliminary study concerning this phenomenon using two large data sets (Lyzenga and Rhebergen, 2007) showed the counter intuitive effect of higher SII values in quiet for normally hearing listeners as compared to hearing-impaired listeners. There is no simple explanation for these differences in predicted speech intelligibility. Both in noise and in quiet, the SII model calculates the effective proportion of speech available to the listener. In stationary noise, the SRTn is generally measured well above the listeners' hearing threshold; in most studies the noise level is between 60 and 80 dBA. The average observed SRTn (i.e., the SNR for speech) is about −5 dB for normally hearing listeners (Plomp and Mimpen, 1979). So, for SRT measurements in stationary noise, the speech levels usually vary from about 55 to about 75 dBA. In quiet, the SRTq is expressed in speech level (dBA), and is highly dependent on the listeners' hearing threshold. The speech level at the SRTq is about 15–20 dBA for normally hearing listeners (for a Dutch corpus, see Plomp and Mimpen, 1979; for a review, see Plomp, 1986). Because in these studies the SRT always represents 50% correct performance, a performance intensity (PI) function (e.g., Sherbecoe and Studebaker, 1990, 2002; Dubno *et al.*, 2002) is not needed to transform SII values and speech levels (or SNRs) into performance correct scores, or visa versa. Therefore, it is not likely that the variance in the SII is due to a flaw in a PI function, it is probably the result of a shortcoming of the SII model calculation.

In general, all of the articulation schemes [i.e., AI, SII, and the speech transmission index (STI) of Houtgast and Steeneken, 1985] were designed to predict speech intelligibility for speech presented at a normal speech-spectrum level (62.35 dB SPL, ANSI S3.5-1997, 1997) to normally hearing listeners. Many listeners with sensorineural hearing impairment not only exhibit increased absolute thresholds (which all articulation schemes take into account), but also suffer from a deterioration of supra-threshold sounds and speech processing capabilities at higher speech levels (Pavlovic, 1989). Pavlovic *et al.* (1986), Studebaker *et al.* (1997), and Ching *et al.* (2001) developed a method to incorporate this supra-threshold deterioration into the AI model, and in the SII (ANSI S3.5-1997, 1997) it has been taken into account by means of a general "hearing loss desensitization" factor. This addition to the articulation theory improved the speech intelligibility predictions for the *average* mild to moderate hearing-impaired listener at higher speech levels (>73 dB

SPL), but it is still difficult to accurately predict the performance of individual hearing-impaired listeners (Ching *et al.*, 1998, 2001; Hogan and Turner, 1998; Studebaker *et al.*, 1999b; Sherbecoe and Studebaker, 2003). Since this hearing loss desensitization only comes into play for large hearing losses, and the level distortion factor for relatively high speech levels, SII calculations for SRTq data (<73 dB SPL), measured in listeners with normal and moderate impaired hearing, will hardly be influenced by this correction. Next to raised thresholds, no further auditory differences between normally hearing and hearing-impaired listeners, such as a loss of compression, are taken into account in the SII.

Contrary to the standardized loudness model (ANSI S3.4-2007, 2007), the SII model does not incorporate an auditory filterbank in its calculations to account for hearing loss dependent changes in spectral resolution. The addition of the upward spread of masking function proposed by Ludvigsen (1987) in the SII model improved to some extent the effect of spectral masking in normally hearing listeners, but is still not sufficient to account for masking as a function of frequency (Noordhoek *et al.*, 1999), as a function of level (e.g., Studebaker *et al.*, 1999), or as a function of hearing impairment (Noordhoek *et al.*, 2000; Hornsby and Ricketts, 2006). Furthermore, the SII model does not account for changes in cochlear compression as a function of level and hearing loss. In the SII, the effective proportion of audible speech is calculated with the aid of an intensity importance function (IIF). In the IIF, the speech dynamic range is assumed to be fixed at 30 dB for all speech levels. Since cochlear compression influences the dynamics of perceived sounds, and is much larger near normal speech levels than near threshold (Glasberg and Moore, 2000; Oxenham and Bacon, 2003; Oxenham and Plack, 1997; Plack and Oxenham, 1998; Plack *et al.*, 2007), it presumably affects the internal dynamics of speech signals as well. As a result, for normally hearing listeners, the internal effective dynamic range of speech presented at a normal speech level is more compressed (smaller) than at relatively low levels. This effect, and its possible interactions with the absolute threshold, are not included in the SII model. Compared to listeners with normal hearing, who may have strong cochlear compression at speech levels (about 65 dB SPL), the effective internal dynamic range of the speech signal for listeners with cochlear hearing loss is less compressed at such levels (giving a larger speech dynamics, i.e., larger vowel to consonant ratios). At lower levels, normally hearing listeners have less compression, which reduces the differences in speech dynamic range between these two groups. Since the articulation theory assumes that the proportion of speech (i.e., audibility for speech presented in quiet) is related to the dynamics of the speech signal, the SII model may be improved by introducing a cochlear compression parameter (in combination with an auditory filterbank: ANSI S3.4-2007, 2007) in the calculations. In a sense, we introduce compression in the present model by introducing cochlear amplification. Cochlear amplification (gain) is highest and approximately linear at very low levels (roughly below 30-dB SPL, e.g., Oxenham, 1995; Glasberg and Moore, 2000). It diminishes for intermediate levels, which result in compression when its effectiveness diminishes and eventually the transduction returns to linear (no gain) for very high levels (roughly higher than 100 dB SPL). Varying this compression parameter (i.e., gain) as a function of hearing loss will affect the internal representation of the speech effective dynamic range (i.e., the IIF). Therefore, the calculation scheme is non-linear for normally hearing listeners and becomes more and more linear with increasing hearing loss. Because of cochlear compression, the amount of audible speech may change less as a function of input level for normally hearing listeners as for hearing-impaired listeners. As a result the SII variance at the SRT in quiet may be reduced for normally hearing listeners. Furthermore, the differences between the SII values in quiet and noise might be smaller when accounting for the IIF as a function of level and hearing loss. In addition, reducing the compression when modeling hearing-impaired listeners will give rise to a more realistic model. Both these adaptations may increase the predictive power of the model and to lead to a better description of SRTq data for normally hearing and hearing-impaired listeners.

This paper describes a novel speech intelligibility approach to model SRTs for sentences presented in quiet and in noise, for normally hearing and hearing-impaired listeners. With the aid of the ANSI loudness model (ANSI S3.4-2007, 2007; Moore and Glasberg, 2004) and a cochlear compression function (Oxenham, 1995), expected excitation levels are calculated for speech and for noise, in which the applied compression function is varied with level and hearing loss. First, the assumed dynamic range of speech is converted to excitation patterns. Next, SII values for presented speech levels are calculated by comparing those levels with the excitation levels demarking the dynamic range of speech, the absolute threshold, and, when present, the noise. The inclusion of compression in the SII model is expected to decrease the observed differences between estimated SII values for SRTs in quiet and in stationary noise, especially for normally hearing listeners (see Sec. II B for how introducing compression in the model might alleviate this problem). In addition, if it is a better model, the variance in SII values within and between groups of normally hearing and hearing-impaired listeners should decrease. In addition to compression, there are other factors, both cochlear and central, that can affect speech intelligibility. In the present approach we start by incorporating just cochlear compression in the model, which may be extended with more cochlear factors in later studies (e.g., forward masking). When all cochlear factors are accounted for, investigations concerning more central aspects of speech perception, such as across-channel interactions and phonemic restoration, can be considered.

II. MODEL DESCRIPTIONS

A. The standard SII model

A detailed description of the SII model is given in the standard ANSI S3.5-1997, 1997 or alternatively in Amlani *et al.*, 2002. In short, the SII model calculates the average proportion of speech information available to a listener. As input, it uses the long-term average spectra of the presented speech and noise, and the listener's pure-tone hearing levels.

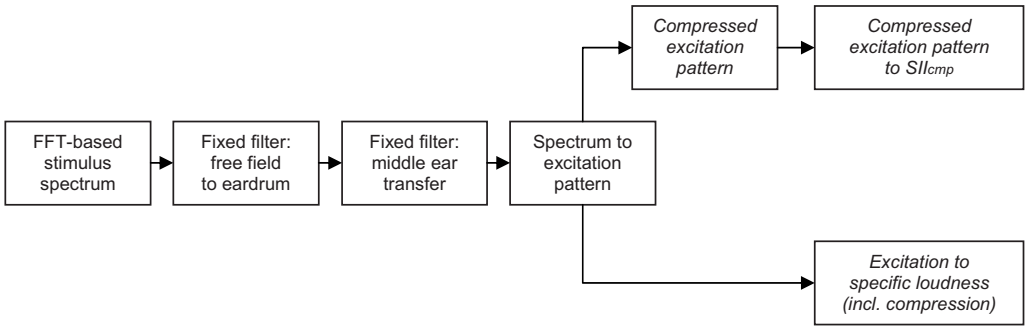


FIG. 1. Schematics of the models. The loudness model of Moore and Glasberg (2004) is depicted in the lower branch. The branching of to the top depicts the SIIcmp model.

Speech and noise are filtered into a number of frequency bands. Within each band, spectrum levels of speech and noise are determined separately, including correction factors for audibility in relation to the absolute threshold for pure tones, for self-masking of speech, and for upward spread of masking. Next, the signal-to-noise ratio is calculated in each band (SNR^i). These SNR^i values are used to calculate the proportion of audible speech (A^i) with aid of an IIF that assumes linearly increasing importance of speech levels from −15 to +15 dB relative to the rms of the speech

$$A^i = (SNR^i + 15)/30;$$

$$\text{when } A^i > 1 \text{ then } A^i = 1,$$

$$\text{when } A^i < 0 \text{ then } A^i = 0. \quad (1)$$

Subsequently, the audibility within each frequency band (A^i) is multiplied with a band-importance function (I^i) to account for the type of speech material (e.g., Studebaker *et al.*, 1993; Pavlovic, 1984, 1987), and with a level distortion factor (D^i) to account for very high speech levels, to yield the effective audibility within a frequency band. Last, the SII is the sum of the effective audibility in each band

$$SII = \sum_i A^i \times I^i \times D^i. \quad (2)$$

B. SIIcmp model with cochlear compression

The present model attempts to alleviate the discrepancy between SII estimates in noise and in quiet by including cochlear compression in the "inner-ear" part of the model. We refer to this model as the SIIcmp model. In the ANSI SII model, compression is not included explicitly in the calculations. However, because this model was designed to work well at normal speech levels (around 65 dB SPL), the effects of compression were included implicitly by using data collected for such levels to calibrate the model. Introducing cochlear compression explicitly into the SIIcmp model allows the possibility of reducing the amount of compression for lower stimulus levels, in accordance with many findings in humans and animal models (Ruggero *et al.*, 1997; Russell and Nilsen, 1997; Rhode and Cooper, 1996). When reducing the compression for low levels, the internal dynamic range of speech at low levels will increase. Estimated excitation levels for the softer speech parts will then benefit less from cochlear amplification, which will reduce their levels, possibly below the absolute threshold. As a consequence, less speech information will be assumed to be available to normally hearing listeners, which will counteract the relatively high standard SII values estimated for those conditions.

The SIIcmp model is based on a series of loudness models for normally hearing and hearing-impaired listeners (Glasberg and Moore, 2002; Moore and Glasberg, 2004; Glasberg and Moore, 2006), which provided the basis for ANSI S3.4-2007, 2007 (middle and lower branches of Fig. 1). Using these models, they were able to predict loudness values for a large variety of conditions and sounds for both normally hearing and hearing-impaired listeners. For the present calculation of SII values, we use the loudness-model stages up to the calculation of the short-term excitation patterns for fluctuating sounds (middle branch of Fig. 1). The model starts with two filter stages to account for the free-field to eardrum sound pressure transformation, and the middle-ear conversion of sound pressure at the eardrum to sound pressure at the oval window of the cochlea. To generate an excitation pattern on an effective equivalent-rectangular bandwidth (ERB) scale, the initial filter stages are FFT (fast Fourier transform) spectral decomposition. To facilitate predictions for fluctuating sounds, a series of short-term excitation patterns is calculated with a 1-ms step, from the consecutive short-term spectra in 64-, 32-, 16-, 8-, 4-, or 2-ms temporal windows, depending on the frequency range of the analysis band (see Table I or Glasberg and Moore, 2002). Normal perceptual sensitivity or spectral resolution is assumed to be accounted for in the calculation of these excitation patterns. In the original model, specific loudness values were calculated next, as a function of frequency, from these un-compressed excitation patterns. In the present

TABLE I. Frequency regions of the FFTs and the corresponding temporal windows.

F_{low} (Hz)	F_{high} (Hz)	Window size (ms)	Window size (samples)
31.25	78.125	64	2048
93.75	468.75	32	1024
500	1 187.5	16	512
1250	2 375	8	256
2500	3 750	4	128
4000	15 000	2	64

model, when calculating the excitation patterns from the 64- to 2-ms short-term spectra, we have explicitly incorporated a compression function (after Oxenham, 1995), as given in the following (left-hand box in the top row of Fig. 1):

$$L < 0: \; L_c = L + G, \tag{3}$$

$$0 <= L < 100: \; L_c = 0.9L + A + B\left(1 - \frac{1}{1 + e^{(50-L)/20}}\right), \tag{4}$$

$$L >= 100: \; L_c = L, \tag{5}$$

where L is the input level, G is the cochlear gain, A is $-0.0894G + 10.894$, and B is $1.1789G - 11.789$. This equation was suggested by Oxenham (1995), and was used subsequently by Moore *et al.* (1996), Oxenham and Moore (1997), Glasberg, *et al.* (1999), Glasberg and Moore (2000), Wojtczak *et al.* (2001), and Bacon and Oxenham (2004). The functions have shapes closely resembling those measured on the basilar membrane in animal models (Robles *et al.*, 1986; Ruggero, 1992; Ruggero *et al.*, 1997).

To produce SII values for the present SRTs in quiet and in stationary noise, we first generated an excitation pattern for the absolute threshold (for normally hearing listeners resulting in excitation levels near 0 dB), and in each temporal window we generated excitation patterns representing the noise (when present) and representing the width of the IIF dynamic range (upper and lower boundaries). By comparing the excitations for the lower and upper boundaries of the speech with the higher one of the internal (absolute threshold) and external noise, signal-above-noise ratios can be calculated. This procedure generates a series of internal signal-above-noise ratio as a function of frequency (ERB scale) and time (1-ms step). In each temporal window, short-term SII values are calculated from these ratios using similar spectral weighting as used in the standard SII. For this step, the spectral weights of the standard SII were recalculated for the ERB scale. At the moment no self-masking of speech is included in these calculation stages. In the standard SII, the self-masking of speech is assumed to lie at a level of 24 dB below the speech rms. This function has only a very limited effect on the estimated SII values, as it falls outside of the dynamic range of speech assumed in the IIF of the ANSI standard. The self-masking of speech in the ANSI SII appears to originate from the earlier version of French and Steinberg (1947) in which the dynamic range of speech was assumed to stretch from 12 dB above rms to 24 dB below. In addition, no level distortion factor is included in the present model, as such a factor is not expected to play a large role in SRTs for sentences in quiet, nor can it be properly calibrated using the data set available for this study (see below). So, at present, any loss of perceptual sensitivity is only modeled in the estimation of the excitation patterns. Finally, the predicted SII is generated by averaging the short-term SII values.

To model hearing impairment, the function of the model is altered in a number of ways. First, the audiogram is assumed to be composed of a degree of inner and a degree of outer-hair-cell loss in a fixed ratio. Moore and Glasberg (2004) found an outer-hair-cell loss versus inner-hair-cell loss ratio of 0.9 to correspond best with their loudness data for hearing-impaired listeners. The outer-hair-cell loss is used to broaden the auditory filters as in Moore and Glasberg, 2004, and to set a loss of cochlear compression by reducing the cochlear amplification G in Eqs. (3)–(5). This reduction in cochlear amplification results in a reduction in excitation level. The inner-hair-cell loss modeled as a further reduction in excitation level. Together, these two reductions in excitation level represent a loss of pure-tone sensitivity in the excitation pattern in accordance with the audiogram.

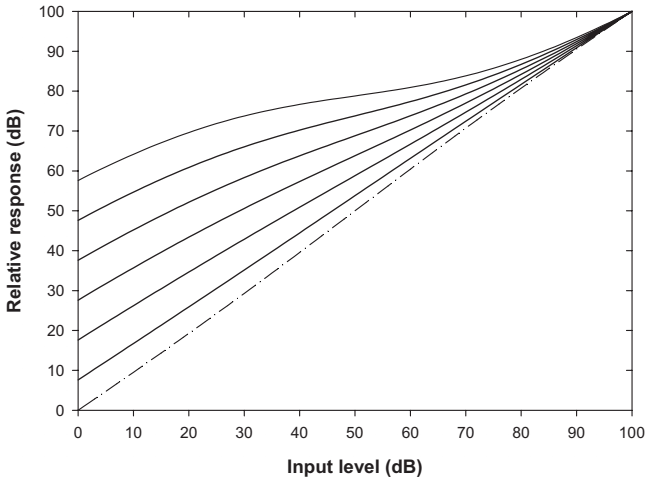


FIG. 2. Schematic input-output cochlear compression function for several values of the cochlear gain G (from bottom to top): 0, 7.6, 17.6, 27.6, 37.6, 47.6, and 57.6 dB.

C. Settings of the SIIcmp model

The SIIcmp model was used to predict SRTs in quiet and in stationary noise for both normally hearing and hearing-impaired listeners for comparison with the standard SII. We used a default outer-hair-cell loss versus inner-hair-cell loss ratio of 0.9. The amount of compression was set from the listeners' audiograms, and the signal levels in the experiments were passed into the model in the same way as done by Moore and Glasberg (2004). The maximum amount of gain G in Eqs. (3)–(5) was set to 57.6 dB (as shown in Fig. 2). For all the calculations, we used stationary speech-shaped noise fragments as stimuli for both the speech and, when present, the noise, as in the ANSI SII model. These fragments were 3661 samples long, sampled at 44.1 kHz, and they were resampled to 2656 samples at a sample rate of 32 kHz (after Moore and Glasberg, 2004). For the present stationary-noise fragments, this resulted in 20 time frames of 64 ms (maximum length of 2048 samples) with a 1-ms (32 samples) step. The resultant 20 short-term SIIcmp values were averaged together to produce a long-term SIIcmp estimate. This procedure was followed to allow modeling of speech reception in fluctuating noise in later studies (for which more time frames will need to be averaged together than for the present stationary stimuli), for the present stationary stimuli it did not affect outcomes.

To model the width of the IIF dynamic range of speech in the excitation patterns, we chose different speech levels and calculated excitation patterns for them. We assume that the effective, perceptual, dynamic range of the speech is defined at the eardrum, and that the dynamic range in the internal representation is compressed depending on presentation level and hearing loss. The SII model assumes a fixed speech dynamic range dynamic range of 30 dB at the eardrum. In a number of studies (Cox *et al.*, 1988; Rankovic, 1997, 1998; Boothroyd, 1990, 2000; van Tasell, 1993; Byrne *et al.*, 1994; Drullman, 1995a; Zeng *et al.*, 2002; Studebaker and Sherbecoe, 2002; Molis and Summers, 2003; Lobdell and Allen, 2007; Hilkhuysen, 2007), dynamic ranges larger than 30 dB were found for a range of speech materials [mean 45 dB, standard deviation (SD)=5 dB across these studies]. To investigate the effects of change in the dynamic range of speech, we used six IIFs with assumed different widths of the IIF dynamic range (30, 35, 40, 45, 50, and 55 dB). The physical speech peaks lie at 12 dB above the rms level (Cox *et al.*, 1988; Byrne *et al.*, 1994). However, we chose the effective speech peaks at 15 dB (Sherbecoe *et al.*, 1993; Studebaker *et al.*, 1999; Sherbecoe and Studebaker, 2002; Studebaker and Sherbecoe, 2002) to account for the notion that listeners are able to listen about 3 dB "into" the noise (Scharf, 1970; Drullman, 1995a, 1995b; Moore, 1997; Studebaker *et al.*, 1999; Sherbecoe and Studebaker, 2002). Thus, the IIFs ranged from +12 dB down to −18, −23, −28, −33, −38, and −43 dB relative to the rms level, respectively. The results for these six different IIFs as function of hearing loss and speech level in quiet and noise were analyzed statistically to help in selecting the appropriate effective IIF dynamic range. This will be discussed further in Sec. VI.

III. METHODS

In a preliminary study (Lyzenga and Rhebergen, 2007), SRT data were used from two studies that measured SRTs for male (Smoorenburg, 1992) or female (Festen and Smits, 2007) speech. In both studies, SRTs were measured in quiet and in noise according to the method of Plomp and Mimpen (1979). An extensive description of this method is described in the review of Plomp (1986). The SRT estimates the SNR at which 50% of presented sentences are repeated entirely correctly. Unlike the Smoorenburg data,[1] the Festen and Smits (2007) data were measured at a single clinic using a constant method, and the speech used was the high-quality female material (Plomp and Mimpen, 1979). Therefore, it was decided to concentrate on that study to evaluate the SII-cmp model settings.

The data set of Festen and Smits (2007) comprises 402 ears. The subjects were all working in airplane maintenance under noisy conditions. Their ages ranged from 18 to 62 years (average of 31 years with a SD of 7). All subjects were native speakers of Dutch. Their tone-audiometric thresholds were recorded at one-third octave frequencies: 0.25, 0.5, 0.8, 1, 1.25, 1.6, 2, 2.5, 3.15, 4, 5, 6.3, and 8 kHz (with the help of continuous audiometry: Peekel D4/6631). The distribution of their thresholds is presented in Fig. 3. In the present paper, the ears were split into three categories: 323 normally hearing denoted as NH, 65 with a dip presumably resulting from noise exposure (a dip of more than 15 dB hearing loss between 2 and 6 kHz relative to 1 and 8 kHz, Smoorenburg, 1992; Bosman and Smoorenburg, 1995) denoted as noise-induced hearing loss (NIHL), and 14 hearing-impaired ears (on average more than 20 dB hearing loss) denoted as HI. Monaural SRTs[2] were measured for the better ear in quiet (SRTq) and in stationary speech-shaped noise (SRTn) presented at four fixed levels: 35, 50, 65, and 80 dBA. Signals were mixed and played out using a PC. The sample rate of all stimuli was 15.625 kHz.

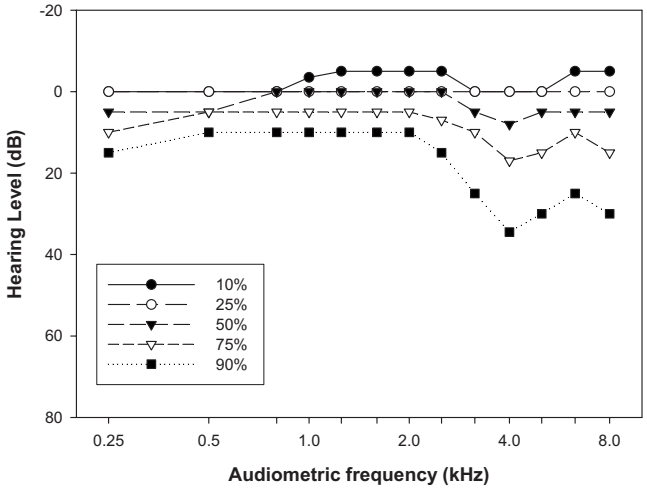


FIG. 3. Distribution of pure-tone thresholds in the set of 402 ears of the Festen and Smits (2007) set. The percentages indicated represent the fraction of ears at a certain frequency with higher tone thresholds than the plotted value.

IV. MODELING RESULTS

A. The data set

Table II shows the distribution of the SRTq (noise level=−∞ dBA) and SRTn (in dB SNR) values, for all listeners' groups of the Festen and Smits (2007) data set. To examine the effects within the SRTn data set, a four [noise level: 35, 50, 65, and 80 dBA] × three [listener group: NH, NIHL, HI] analysis of variance (ANOVA) was performed. As expected, the effects of "noise level" ($F[3,1596]=98.2$, $p<0.001$) and "listener group" ($F[2,1596]=99.9$, $p<0.001$) were both significant, as was the interaction between them ($F[6,1596]=26.2$, $p<0.001$). Bonferroni *post-hoc* tests showed significant differences between all three listener groups ($p<0.01$), and between the SRTs for all noise levels ($p<0.01$) except for the levels 50 and 65 dBA ($p=0.3$), and the levels 35 and 80 dBA ($p=1.0$). To examine the effects within the SRTq data set, an ANOVA was performed on the three listener groups. As expected, listener group ($F[2,399]=81.5$, $p<0.001$) was significant. Bonferroni *post-hoc* tests showed significant differences between all three listener groups ($p<0.01$). Table III shows the average SRTs for the subgroups.

TABLE II. (a) Distribution of SRTs of the Festen and Smits (2007) data set, measured in quiet, expressed in dBA, and the SRTs measured in noise at 35, 50, 65, and 80 dBA, expressed in dB SNR. Data are shown for the 323 normally hearing listeners in (a), for the 65 listeners with noise-induced hearing losses in (b), and for the 14 hearing-impaired group in (c).

Percentiles	5	10	25	50	75	90	95
SRTq (dBA)	11.8	13.0	15.0	17.4	20.2	24.9	27.3
SRTn (35 dB)	−6.5	−5.8	−5.0	−4.2	−3.0	−2.2	−1.0
SRTn (50 dB)	−7.4	−6.8	−5.8	−5.0	−4.2	−3.4	−3.0
SRTn (65 dB)	−7.0	−6.6	−5.8	−5.0	−4.6	−3.8	−3.4
SRTn (80 dB)	−5.8	−5.4	−4.6	−4.2	−3.4	−2.6	−2.6
SRTq (dBA)	13.0	14.4	16.0	19.2	22.8	27.0	27.8
SRTn (35 dB)	−5.6	−4.8	−4.2	−3.0	−2.0	−1.0	−1.0
SRTn (50 dB)	−7.4	−6.8	−5.8	−5.0	−3.8	−2.4	−1.8
SRTn (65 dB)	−6.9	−6.4	−5.8	−5.0	−4.2	−3.0	−2.3
SRTn (80 dB)	−5.7	−5.0	−4.2	−3.8	−3.0	−2.4	−1.4
SRTq (dBA)	17.4	17.6	25.2	31.6	40.9	52.6	56.2
SRTn (35 dB)	−4.6	−4.4	−2.0	1.2	5.9	16.2	19.0
SRTn (50 dB)	−5.4	−5.4	−4.7	−3.6	−2.4	3.8	8.2
SRTn (65 dB)	−5.4	−5.4	−5	−4.6	−3.3	−1.2	−0.6
SRTn (80 dB)	−5.0	−4.8	−4.6	−4.0	−2.5	−1.8	−1.4

B. The standard SII model

For comparison to the SIIcmp approach with compression, speech intelligibility was also predicted with the standard SII model (ANSI S3.5-1997, 1997). The standard SII calculations were conducted using the long-term speech spectrum of the female speaker. In order to approximate sound levels at the ear drum (as required by the SII model), speech and noises were filtered with a 400th point finite impulse response (FIR) filter with the characteristics of the Beyer Dynamic DT48 headphones used in the study. Absolute thresholds at each 1/3-octave frequency were used as input to the SII model. The standard SII predictions are calculated with a fixed IIF of 30 dB (+15 to −15 dB ref rms) speech dynamic range. For the NH listener group, we assume that no supra-threshold deficits played a role in the recorded SRTn and SRTq values. So, the effective audibility (i.e., the SII) should be similar for SRTs in quiet and in noise. The SII values corresponding to SRTs for listeners with NIHL and HI listeners are often found to be higher (i.e., worse) than those of NH listeners (e.g., Plomp, 1986; Festen and Plomp, 1990; Peters *et al.*, 1998; Versfeld and Dreschler, 2002). This is often attributed to the fact that listeners with cochlear hearing loss need more speech information to reach the same "effective" speech audibility as normally hearing listeners. The SII model does not account for hearing loss in a completely realistic manner, and calculates audibility in stead of effective audibility. As a consequence, the SII values at the SRT may be higher when cochlear impairments are involved.

The boxes in Fig. 4 show the SII values for the three groups of listeners for noise levels of $-\infty$ (quiet), 35, 50, 65, and 80 dBA. In each box triplet, the boxes on the left, in the middle, and on the right represent the groups NH (n=323), NIHL (n=65), and HI (n=14), respectively. As expected, the ANSI S3.5 predicts different SIIs for speech intelligibility in quiet and in noise: For NH listeners, the average predicted SII in quiet (0.39) is about 0.10 higher than the SII for a stationary-noise level of 80 dBA (0.29). In quiet and for low noise levels (35 and 50 dBA), the SII values are highest for the NH listeners, followed by the NIHL listeners, and lowest for the HI listeners. These SII results suggest that normally hearing rather than listeners with a cochlear deficit have supra-threshold deficits in quiet and in low noise conditions. In higher noise conditions (65 and 80 dBA), the SII values for the three listeners' groups are rather similar and show no consistent trends. When concentrating on the spread in the data, it can be seen that the variance in the SIIs in quiet is relatively large (NH: 0.011), even though we used individual 1/3-octave audiograms, and that the spread goes down with increasing stationary-noise levels (NH: from 0.003 at 35 dBA down to 0.001 at 80 dBA).

C. The SIIcmp model

For the SIIcmp approach, SIIs' predictions are calculated assuming six different widths of dynamic range of speech (IIFs stretching over 30, 35, 40, 45, 50, and 55 dB,

TABLE III. SRTs measured in the quiet, expressed in dBA, and the SRTs measured in noise at 35, 50, 65, and 80 dBA, expressed in dB SNR for NH, NIHL, and HI. Standard deviations are given between parentheses.

	NH	(n=323)	NIHL	(n=65)	HI	(n=14)
SRTq (dBA)	17.5	(4.1)	19.6	(4.9)	33.4	(11.3)
SRTn (dB 35)	−4.0	(1.7)	−3.1	(1.5)	3	(6.8)
SRTn (dB 50)	−5.2	(1.3)	−4.7	(1.7)	−2.8	(3.5)
SRTn (dB 65)	−5.2	(1.2)	−4.8	(1.3)	−4	(1.4)
SRTn (dB 80)	−4.1	(1)	−3.7	(1.1)	−3.6	(1.1)

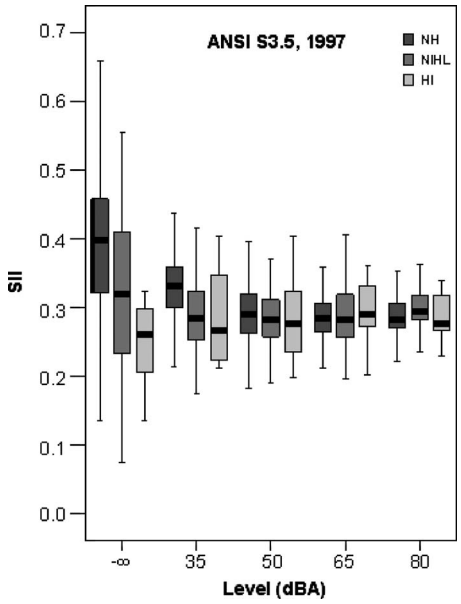


FIG. 4. Predicted ANSI SII values for the Festen and Smits (2007) data set. In each box triplet, the boxes on the left, in the middle, and on the right represent the groups NH, NIHL, and HI, respectively. From left to right, the triplets show the predictions for the SRTq and for four SRTns at noise levels of 35, 50, 65, and 80 dBA, respectively.

respectively). The speech peaks are assumed to be 12 dB for all six IIFs, representing the physical peaks above the rms. The box triplets in the six panels of Fig. 5 show the resulting of six sets of SII predictions, from left to right in each panel for the SRTs in quiet ($-\infty$) and in stationary noises at 35, 50, 65, and 80 dBA, respectively. In quiet, the SIIcmps show a large effect of the chosen IIF dynamic range, for the NH group (left-most boxes of each triplet) the predictions go down from 0.40 at a 30-dB range to 0.18 at a 55-dB range, while their variances decrease from 0.011 to 0.002 with increased IIF dynamic range. So, depending on the chosen dynamic range, the present model improves intelligibility predictions in comparison to the standard SII. In stationary noise, the predictions go down from around 0.3 at the 30-dB range to just under 0.2 at the 55-dB range, while their variances are very small (between 0.001 and 0.002) for all IIF dynamic ranges. It is interesting to compare the SIIcmp results in Fig. 5 to the standard SII results in Fig. 4. In Fig. 4, it can be seen for quiet and for the 35-dBA noise level that the amount of speech needed for comprehension is smaller for larger hearing losses. This is inconsistent with the idea that hearing-impaired listeners have supra-threshold deficits further reducing intelligibility, and it is not seen in the panels of Fig. 5 for the SIIcmp. Figure 5 shows for all three groups that the variances in the predictions are much higher for the SRTq than for the SRTn, which may well be related to relatively high uncertainty in the audiogram data. For all three groups, the predicted SIIcmps for the SRTq and the SRTn are closest in value for the IIF dynamic range of 45 dB. For this dynamic range, the variances for the SRTq are smaller for the SIIcmp (NH: 0.004; NIHL: 0.005; HI: 0.010) than for the standard SII prediction (NH: 0.011; NIHL: 0.015; HI: 0.017). This is illustrated in the bottom panel of Fig. 6, which shows the average SII values in noise per listener group compared to the average SII in quiet for the ANSI SII and the SIIcmp for the chosen IIF dynamic ranges. The average SII in quiet lies within the spread of the SII values in noise for dynamic ranges between about 45 and 50 dB. In Fig. 6, the left-most values are those for the ANSI SII. The top panel of Fig. 6 illustrates the reduced spread between groups for the SIIcmp in quiet compared to the ANSI SII in quiet. When the dynamic range in the standard SII calculation was increased in a similar manner as was done for the SIIcmp, then the SII values decreased, but the relative distance between the expectations in quiet and in noise did not decrease (not shown in Fig. 6), as can be expected for a linear model.

D. Analysis of variance

To test whether the noise level affected the audibility calculations, a 4 [noise level(35,50,65, and 80 dBA] $\times 3$ [group: NH, NIHL, and HI] multivariate analysis of variance (MANOVA) was performed on the 7 SII data sets [SII model calculations: SIIcmp for 6 IIF dynamic ranges and the standard SII] for listeners shown in Figs. 4 and 5. Of the main effects, noise level and group were significant. This indicates that the dynamic models were unable to produce consistent SII values as function of noise level and listeners' group. However, the interaction noise level $\times$ group was only significant for the ANSI SII predictions ($F[6,1596]=7.2$, $p<0.0001$). This indicates that the ANSI S3.5 model was not able to produce consistent predictions for the SRTn across noise levels and groups, while the SIIcmp approach managed to do so. As noted before, the ANSI SII values for lower noise levels are higher for NH compared to NIHL and HI listeners, while the opposite is true for higher noise levels. This may be explained by interactions, between the varying amounts of compression and the absolute threshold that can occur for the lower noise levels in the SIIcmp calculations.

One of the aims of our investigation was to bring predicted SII values for different hearing losses and for conditions in quiet and in noise closer together than they are for the standard SII. To test this, a 2 [listening condition: quiet and pooled noise conditions] $\times 3$ [listener group] MANOVA was performed on the 7 SII data sets [SII model calculations: SIIcmp for 6 IIF dynamic ranges and the standard SII]. Of the main effects, the effect of group was significant only for the ANSI SII predictions ($F[2,2004]=47.9$, $p<0.0001$). The main effect of "listening condition" was only significant for the ANSI SII and for the SIIcmp at 30–40 dB IIF dynamic ranges. So, for larger dynamic ranges, the SIIcmp showed no significant effects of group. In addition, no significant interactions were found for the SIIcmp results for IIF dynamic ranges of 35 and 40 dB. However, the articulation theory assumes that the presentation level has no effect on the speech intelligibility around normal speech levels, whereas for higher levels a level distortion effect occurs (>73 dB SPL). When the 80 dBA noise level is excluded from the pooled SRTn data, to eliminate any possible effects of level distortions, a significant main effect for the factor group is still only shown for the ANSI SII calculations ($F[2,1604]=42.7$, $p<0.0001$). The main effect of listening condition was only significant for the ANSI SII and the SIIcmp for 30-

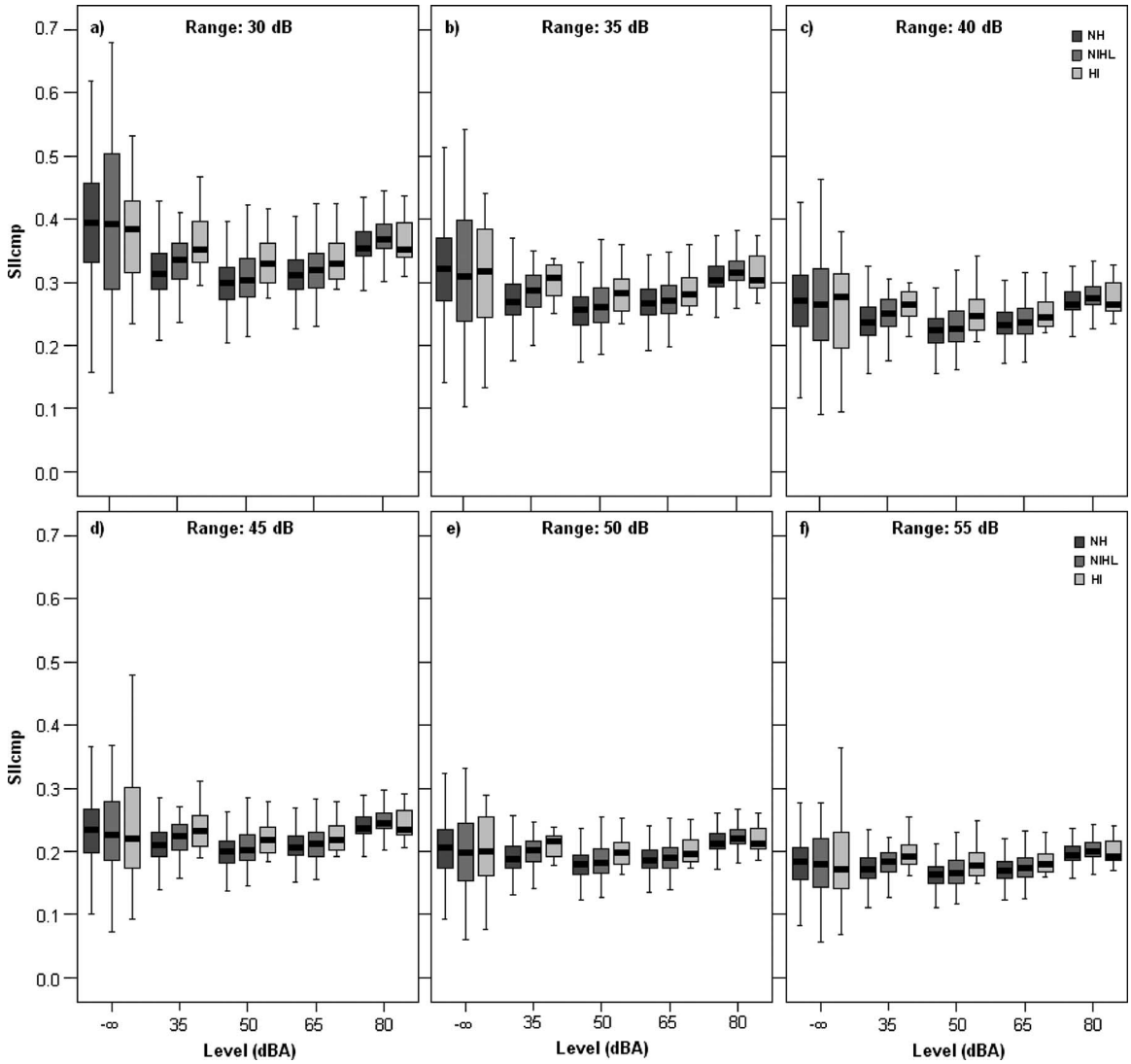


FIG. 5. Predicted SIIcmp values for the Festen and Smits (2007) data set for the chosen IFF dynamic ranges of 30, 35, 40, 45, 50, and 55 dB. The layout of each panel is the same as in Fig. 4.

to 45-dB IIF dynamic ranges. No significant interactions were found for the SIIcmp calculated for an IIF dynamic range of 40 dB.

E. The role of supra-threshold problems in hearing-impaired listeners

For hearing-impaired listeners, SII values above 0.33 at the SRT have often been interpreted as indication of supra-threshold problems (e.g., Noordhoek *et al.*, 2000, 2001; van Schijndel *et al.*, 2001; George *et al.*, 2006). Correcting for their loss of compression may be expected to affect the corresponding speech intelligibility estimations, and therefore could influence the estimated magnitudes of the supra-threshold problems. Since soft sounds will become even harder to perceive without cochlear amplification, it can be expected that including cochlear amplification and compression into the SII calculations will reduce the estimated degree of supra-threshold problems. The model predictions in Secs. IV B and IV C were conducted on a large group of NH subjects (323), but only for relatively small groups of NIHL subjects (64) and HI subjects (14). To further examine the responses of hearing-impaired listeners, SIIcmp values were predicted for SRTs measured in several other studies in literature, with ears classified as NH, NIHL, or HI using the same criteria as were used for the Festen and Smits (2007) data set. Additional data came from the following studies: from Smoorenburg (1992): 400 ears classified into 228 NH, 130 NIHI, and 42 HI; from de Laat and Plomp (1983): 10 young NH and 20 young HI listeners; from Festen and Plomp (1990): 20 HI listeners; and from Bosman and Smoorenburg (1995): 19 NIHL (with very similar audiograms compared to the other NIHL groups), 20 NH, and 14 elderly HI listeners. Female speech (Plomp and Mimpen, 1979) was used to measure the monaural SRT, and all subjects were tested in a sound treated booth with headphones in all studies but Smoorenburg (1992), who used male speech and different measurement rooms not all of which were sound treated. Figure 7 shows the distribution of the standard SII and SIIcmp predictions derived from the data of these studies. For the hearing-impaired listener groups, this figure does not show a clear decrease in the estimates for the SIIcmp as compared to the standard SII. So, for the present sets

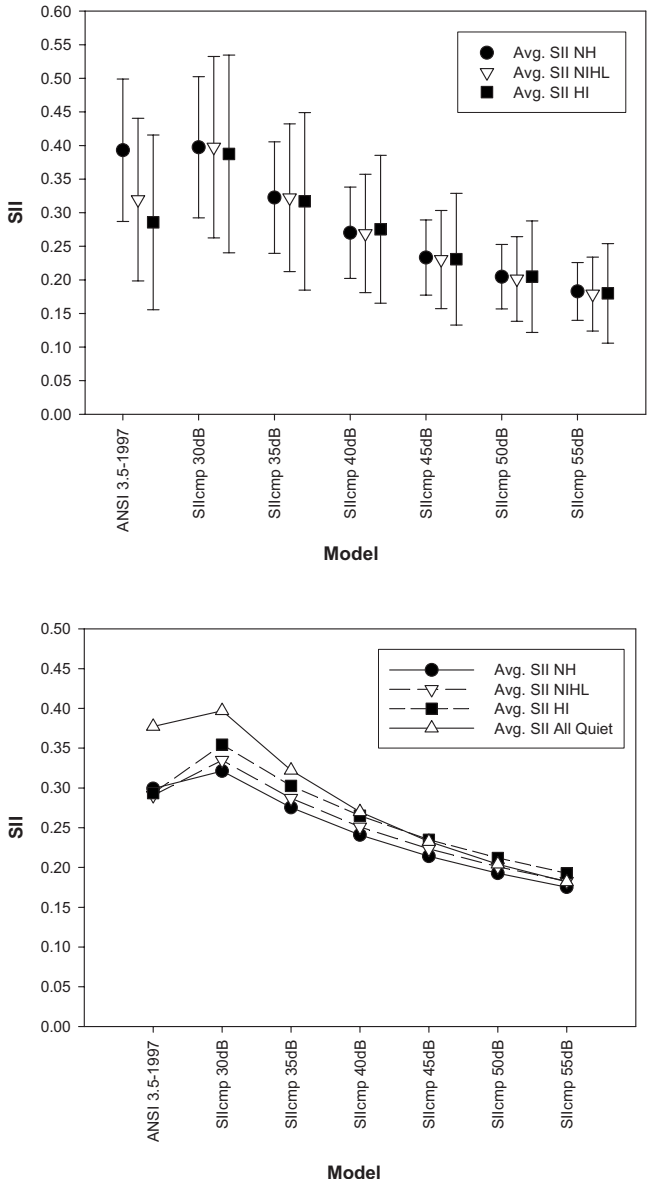


FIG. 6. The top panel shows the average SII values as a function of model (ANSI SII and SIIcmp for the chosen six IFF dynamic ranges) for the SRTq in NH, NIHL, and HI listeners. The results for the listener groups NH, NIHL, and HI are displayed as circles, triangles, and squares, respectively. Error bars indicate the standard deviations between subjects. The bottom panel shows for the corresponding SII values for the SRTn, supplemented with the overall averaged results for the listener groups in quiet (up-side down triangles).

of data, the aforementioned hypothesis of reduced measures of supra-threshold problems by including auditory compression in the speech intelligibility estimations is not confirmed. A more decisive test of this hypothesis will probably require audiograms taken at 1/3-octave frequencies rather than the octave frequencies used in the studies included above.

V. GENERAL DISCUSSION

A. SII predictions

Model predictions with the SIIcmp approach show a significant improvement for the NH and NIHL listener groups of the Festen and Smits (2007) data set. The SIIs for the SRT in quiet and stationary noise are about the same with assumed IIF dynamic ranges of 35, 40, and 45 dB. Furthermore, for the SRTq data, the variance of the predicted SIIcmps is smaller than that of the predicted ANSI SIIs. So, the predictive power for individual SRTs has improved by introducing cochlear compression into the SII model. This may be somewhat surprising, as cochlear compression, or rather cochlear amplification, brings many benefits. Nevertheless, it is not always advantageous. For example, the detection of level increments shows a "mid-level hump" where performance is worse than at other levels (e.g., Carlyon and Moore, 1984). This is being attributed to compression. In addition, there is a body of work arguing that compression reduces modulations, which may be detrimental for speech reception (e.g., Plomp, 1988). In the present paper, we introduced compression into the SIIcmp model, and for the SRTq it pushed a portion of the speech signal below the absolute threshold. So here, it is primarily an interaction between the amount of cochlear compression as a function of level and the absolute threshold that affects SII values for speech perception measurements.

The modified SII approach shows much less improvement for listeners with moderate to severe hearing loss than for normally hearing listeners. It is likely that other factors than audibility have an effect on the speech intelligibility for these listeners. The ANSI SII and the SIIcmp models account for different aspects of hearing loss, and so leave different "remainders" that might be classified as supra-threshold deficits. The SIIcmp model attempts to account for cochlear compression as a function of hearing loss, but it does not account for other factors such as degraded temporal resolution. Furthermore, while the model does take some cochlear factors into account, it does not account for factors that have more central causes (e.g., cognition deficits), or factors related to Tinitus, Hyperacusis, or Menière's disease. Such problem can affect the robustness of the measured speech reception levels. Therefore, model predictions for these types of hearing loss are likely to be less reliable than those for listeners with a more simple cochlear hearing loss. To account for other factors that may play a role in speech intelligibility, one should not just rely on the listener's hearing levels. Additional aspects of the listeners' performance (i.e., a full auditory profile) might also need to be taken into account (Dreschler *et al.*, 2007).

B. Effective speech dynamic range

At normal speech levels, it can be assumed that cochlear compression will "squeeze" the speech dynamic range together in a more or less uniform way. So, adding compression to the SII model does not directly change the estimated SII values, as they are calculated from ratios of available speech information in the dynamic range. That is, unless there is a knee-point of the compression function in the dynamic range of the speech, or unless there is an interaction with the absolute threshold. The effects of compression for the SRTq in the current study probably arise from an interaction between compression and absolute thresholds. In a previous study presented at the ASA meeting in Salt Lake City (Lyzenga and Rhebergen, 2007) we have looked into the

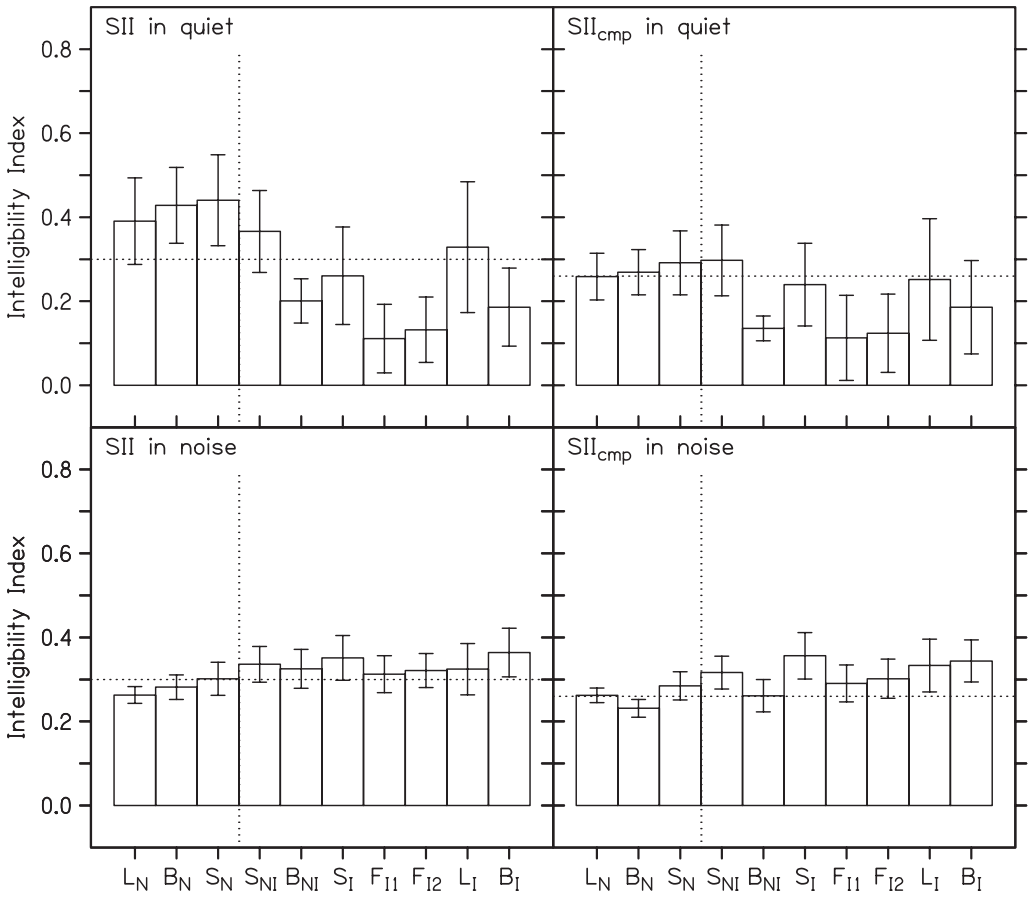


FIG. 7. Predicted SII and SIIcmp values, for an IIF dynamic range of 40 dB, for the data from Smoorenburg (1992), denoted as S_N, S_{NI}, and S_I; from Bosman and Smoorenburg (1995), denoted as B_N, B_{NI}, and B_I; from de Laat and Plomp (1983), denoted as L_N and L_I; and for the set of 20 hearing-impaired ears of the data from Festen and Plomp (1990), denoted as F_{I1} and F_{I2} for two types of used speech material, where the subscripts N, NI, and I stand for normal hearing (to the left of the vertical dotted line), noise-induced hearing loss, and hearing impaired, respectively. Note that the average values for the SII and the SIIcmp are not the same (horizontal dotted lines).

difference in SIIcmp predictions between the following two assumptions: (1) The effective (perceptual) dynamic range of the speech is 30 dB, is constant at the eardrum, and is compressed depending on presentation level and hearing loss, and (2) the effective dynamic range of the speech is fixed at 15 dB in the excitation (after compression with a ratio of 2 at the "normal" speech level). The second assumption did not lead to any systematic and understandable results. So, in the present study, we dropped that assumption and continued on with the assumption that the dynamic range of speech is defined at the eardrum and is affected by cochlear compression before it expresses itself in the excitation.

The effective speech dynamic range is assumed to be 30 dB in the SII model (ANSI S3.5-1997, 1997), with the speech peaks at 15 dB above the rms. The latter assumption is based on a range of speech perception experiments (e.g., Studebaker *et al.*, 1993, 1999; Studebaker and Sherbecoe, 2002). The 30-dB effective speech dynamic range is based on a "best fit" of the data of Studebaker *et al.*, 1993 (Sherbecoe and Studebaker, 2002). Recently, some studies have been published where the effective speech dynamic range was measured in intelligibility experiments using words (Studebaker *et al.*, 1999; Sherbecoe and Studebaker, 2002) or sentences (Drullman, 1995a; Hilkhuysen, 2007). The results of these studies suggest that the effective dynamic range of the speech is likely to be larger than the originally assumed 30 dB, and it is probably closer to a value between 40 and 45 dB. Based on these perceptual experiments, model predictions with such assumed dynamic ranges offer plausible alternatives. In the present study, SIIcmp predictions for larger assumed dynamic ranges (>30 dB) reduced the variance in the predictions for the SRTq, whereas the variances in SIIcmp predictions for the SRTn data remain fixed at very low values. In contrast, the predictions of the standard SII did not improve when the dynamic range in that model was increased (Sec. IV C). The effects of hearing threshold are clearly smaller for the SRTn measurements, which were performed at relatively high presentation levels. Under such conditions, the noise level is well above the SRTq ($\geq$20 dB), and consequently the noise level will be the main factor limiting speech audibility.

For lower speech levels, and for listeners with hearing losses at levels near the noise levels, the audibility of the speech will be dominated by the either the noise or the absolute threshold, whichever has the higher level. For the SRTq, intelligibility primarily depends on the speech levels relative to the absolute threshold, which is reflected in the SII models. Since the absolute thresholds are often measured at octave frequencies, and with an accuracy of about $\pm$ 5 dB,

the reliability of the absolute-threshold estimate is a major factor in the variance of SII predictions in quiet. A 5-dB error per octave frequency may lead to approximately a 16.7% audibility difference compared to an ideal prediction. As a consequence, the variance in SII predictions can be expected to be larger for SRTq data than for SRTn data. The advantage of the Festen and Smits (2007) data set was that hearing thresholds were measured at 1/3-octave intervals, which gives a better spectral sampling of the actual absolute-threshold profile than octave frequencies. Using absolute thresholds recorded at 1/3-octave frequencies would likely lead to more reliable results than using just the octave frequencies for which the audiogram is routinely measured.

VI. CONCLUSIONS

Introducing compression in speech intelligibility calculations increased the predictive power of the model: For normally hearing listeners, the SII values for SRTs in quiet and in stationary noise were brought closer together and the spread in the data was reduced. This can probably be attributed to interaction effects between compression and absolute thresholds. For hearing-impaired listeners no clear decrease in estimated supra-threshold deficits was observed. In the future, more cochlear effects such as forward masking will need to be included in the model. When the function of the cochlea is modeled with greater precision, the amount of supra-threshold deficits in hearing-impaired listeners might need to be re-evaluated.

ACKNOWLEDGMENTS

K.S.R. was financially supported by the Dutch Foundation "Heinsius Houbolt Fonds." The authors like to thank Guido Smoorenburg, Arjan Bosman, and Jan de Laat for making their SRT data sets available to them. They also want to thank Brian Moore for the MATLAB® code of the loudness model. The constructive comments of the two anonymous reviewers and the editor Ken Grant are gratefully acknowledged.

[1]The Smoorenburg (1992) data were collected at five different clinics with slightly different methods, and they show an effect of location.

[2]The speech reception threshold should not be confused with the speech recognition threshold, which is defined as the level at which 50% of spondee words are repeated correctly.

Allen, J. B. (**1996**). "Harvey Fletcher's role in the creation of communication acoustics," J. Acoust. Soc. Am. **99**, 1825–1839.

Amlani, A. M., Punch, J. L., and Ching, T. Y. C. (**2002**). "Methods and applications of the audibility index in hearing aid selection and fitting," Trends Amplif. **6**, 81–129.

ANSI S3.4-2007 (**2007**). "Procedure for the computation of loudness of steady sounds" (ANSI, New York).

ANSI S3.5 (**1969**). "American national standard methods for the calculation of the articulation index" (ANSI, New York).

ANSI S3.5-1997 (**1997**). "Methods for the calculation of the speech intelligibility index" (ANSI, New York).

Bacon, S. P., and Oxenham, A. J. (**2004**). "Psychophysical manifestations of compression: Hearing-impaired listeners," in *Auditory Compression*, edited by S. P. Bacon, A. N. Popper, and R. R. Fay (Springer, New York), pp. 107–152.

Boothroyd, A. (**1990**). "Articulation index: Importance function in the intensity domain," J. Acoust. Soc. Am. **88**, S31.

Boothroyd, A. (**2000**). "Thar's gold in them thar hills: Mining the P/I function Carhart Memorial," lecture delivered to the Annual Convention of the American Audiology Society, Scottsdale, AZ.

Bosman, A. J., and Smoorenburg, G. F. (**1995**). "Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment," Audiology **34**, 260–284.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Heto, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westermann, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Carlyon, R. P., and Moore, B. C. J. (**1984**). "Intensity discrimination: A severe departure from Webers law," J. Acoust. Soc. Am. **76**, 1369–1376.

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Ching, T. Y. C., Dillon, H., Katsch, R., and Byrne, D. (**2001**). "Maximising effective audibility in hearing aid fitting," Ear Hear. **22**, 212–224.

Cox, R. M., Matesich, J. S., and Moore, J. N. (**1988**). "Distributions of short-term rms levels in conversational speech," J. Acoust. Soc. Am. **84**, 1100–1104.

de Laat, J. A. P. M., and Plomp, R. (**1983**). "The reception threshold of interrupted speech for hearing-impaired listeners," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer, Berlin), pp. 359–363.

Dreschler, W. A., Esch Van, T. E., Larsby, B., Hallgren, M., Lutman, M. E., Lyzenga, J., Vormann, M., and Kollmeier, B. (**2008**). "Charactering the individual ear by the 'Auditory Profile' (A)," J. Acoust. Soc. Am. **123**, 3714.

Drullman, R. (**1995a**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Drullman, R. (**1995b**). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," J. Acoust. Soc. Am. **98**, 1796–1798.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Festen, J. M., and Smits, C. (**2007**). "Is speech reception in stationary noise suitable for screening noise induced hearing damage?," Ned. Tijdschr. KNO-Heelk. **13**, 93.

Fletcher, H. (**1953**). *The ASA Edition of Speech and Hearing in Communication*, originally published in 1953, edited by J. B. Allen in 1995 (Acoustical Society of America, Melville, New York).

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–919.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362–2375.

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Glasberg, B. R., and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds," J. Audio Eng. Soc. **50**, 331–342.

Glasberg, B. R., and Moore, B. C. J. (**2006**). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model (L)," J. Acoust. Soc. Am. **120**, 585–588.

Glasberg, B. R., Moore, B. C. J., and Stone, M. A. (**1999**). "Modeling changes in frequency selectivity with level," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier (World Scientific, Singapore).

Hilkhuysen, G. L. M. (**2007**). "Intelligibility trade-off between multi-band peak-limiting and masking by noise," paper presented at the Workshop: Temporal Dynamics in Speech and Hearing, Antwerp, Belgium, 26 August.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.
Holube, I., and Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1716.
Hornsby, B. W. Y., and Ricketts, T. A. (**2003**). "The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding," J. Acoust. Soc. Am. **113**, 1706–1717.
Hornsby, B. W. Y., and Ricketts, T. A. (**2006**). "The effects of hearing loss on the contribution of high- and low frequency speech information to speech understanding. II. Sloping hearing loss," J. Acoust. Soc. Am. **119**, 1752–1763.
Hou, Z., and Thornton, A. R. (**1994**). "A model to evaluate and maximize hearing aid performance by integrating the articulation index across listening conditions," Ear Hear. **15**, 105–112.
Houtgast, T., and Steeneken, H. J. (**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.
Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.
Kryter, K. D. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.
Lobdell, B. E., and Allen, B. A. (**2007**). "A model of the VU (volume-unit) meter, with speech applications," J. Acoust. Soc. Am. **121**, 279–285.
Ludvigsen, C. (**1987**). "Prediction of speech intelligibility for normal hearing and cochlearly hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 1162–1171.
Lyzenga, J., and Rhebergen, K. S. (**2007**). "A model with compression for the estimation of speech intelligibility in quiet," J. Acoust. Soc. Am. **121**, 3199.
Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low-and high-frequency speech," ARLO **4**, 124–128.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, London), pp. 109–111.
Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1996**). "Detection of decrements and increments in sinusoids at high overall levels," J. Acoust. Soc. Am. **99**, 3669–3677.
Müsch, H. (**2001**). "Review and computer implementation of Fletcher and Galt's method of calculating the articulation index," ARLO **2**, 25–30.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners," J. Acoust. Soc. Am. **105**, 2895–2902.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners," J. Acoust. Soc. Am. **107**, 1685–1696.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2001**). "Relations between intelligibility of narrowband speech and auditory functions, both in the 1-kHz frequency region," J. Acoust. Soc. Am. **109**, 1197–1212.
Oxenham, A. J. (**1995**). "Psychophysical consequences of peripheral nonlinearity," unpublished Ph.D. thesis, University of Cambridge, Cambridge, UK.
Oxenham, A. J., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352–366.
Oxenham, A. J., and Moore, B. C. J. (**1997**). "Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ).
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.
Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.
Pavlovic, C. V. (**1989**). "Speech spectrum considerations and speech intelligibility predictions in hearing aid evaluations," J. Speech Hear. Res. **54**, 3–8.
Pavlovic, C. V. (**1994**). "Band importance functions for audiological applications," Ear Hear. **15**, 100–104.
Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50–57.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Plack, C. J., Carcagno, S., and Oxenham, A. J. (**2007**). "A further test of the linearity of temporal summation in forward masking," J. Acoust. Soc. Am. **122**, 1880–1883.
Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.
Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception-threshold of the hearing-impaired," J. Speech Hear. Res. **29**, 146–154.
Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am. **83**, 2322–2327.
Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.
Rankovic, C. M. (**1997**). "Prediction of speech reception for listeners with sensorineural hearing loss," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ).
Rankovic, C. M. (**1998**). "Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **103**, 1043–1057.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2006**). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. **120**, 3988–3997.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2008**). "Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing," Ear Hear. **29**, 169–175.
Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apex of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A. (**1992**). "Responses to sound of the basilar membrane of the mammalian cochlea," Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.
Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, San Diego, CA), pp. 159–202.
Sherbecoe, R. L., and Studebaker, G. A. (**2002**). "Audibility-index functions for the connected speech test," Ear Hear. **23**, 385–398.
Sherbecoe, R. L., and Studebaker, G. A. (**2003**). "Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the Connected speech test," Ear Hear. **24**, 71–88.
Sherbecoe, R. L., and Studebaker, G. A. (**1990**). "Regression equations for the transfer functions of ANSI S3.5-1969," J. Acoust. Soc. Am. **88**, 2482–2483.
Sherbecoe, R. L., Studebaker, G. A., and Crawford, M. R. (**1993**). "Speech spectra for six recorded monosyllabic word tests," Ear Hear. **14**, 104–111.
Smoorenburg, G. F. (**1992**). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–437.
Studebaker, G. A., Gray, G. A., and Branch, W. E. (**1999b**). "Prediction and statistical evaluation of speech recognition test scores," J. Am. Acad. Audiol. **10**, 355–370.
Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.
Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993**). "Frequency importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech Hear. Res. **36**, 799–807.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gray, G. A.

(**1997**). "Age-related changes in monosyllabic word recognition performance when audibility is held constant," J. Am. Acad. Audiol. **8**, 150–162.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.
van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**2001**). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 529–542.
van Tasell, D. J. (**1993**). "Hearing loss, speech, and hearing aids," J. Speech Lang. Hear. Res. **36**, 228–244.
Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.
Wojtczak, M., Schroder, A. C., Kong, Y. Y., and Nelson, D. A. (**2001**). "The effect of basilar membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing," J. Acoust. Soc. Am. **109**, 1571–1586.
Woods, W. S., Van Tassel, D. J., Ricket, M. E., and Trine, T. D. (**2006**). "SII and fit-to-target analysis of compression system performance as a function of number of compression channels," Int. J. Audiol. **45**, 630–644.
Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). "Speech dynamic range and its effect on cochlear implant performance," J. Acoust. Soc. Am. **111**, 377–386.

A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent

Jean-Luc Schwartz[a)]
Department of Speech and Cognition/Institut de la Communication Parlée, GIPSA-Lab, UMR 5216, CNRS, Grenoble University, 38402 Saint Martin d'Hères Cedex, France

(Received 21 November 2008; revised 21 December 2009; accepted 22 December 2009)

Audiovisual perception of conflicting stimuli displays a large level of intersubject variability, generally larger than pure auditory or visual data. However, it is not clear whether this actually reflects differences in integration per se or just the consequence of slight differences in unisensory perception. It is argued that the debate has been blurred by methodological problems in the analysis of experimental data, particularly when using the fuzzy-logical model of perception (FLMP) [Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Laurence Erlbaum Associates, London)] shown to display overfitting abilities with McGurk stimuli [Schwartz, J. L. (2006). J. Acoust. Soc. Am. **120**, 1795–1798]. A large corpus of McGurk data is reanalyzed, using a methodology based on (1) comparison of FLMP and a variant with subject-dependent weights of the auditory and visual inputs in the fusion process, weighted FLMP (WFLMP); (2) use of a Bayesian selection model criterion instead of a root mean square error fit in model assessment; and (3) systematic exploration of the number of useful parameters in the models to compare, attempting to discard poorly explicative parameters. It is shown that WFLMP performs significantly better than FLMP, suggesting that audiovisual fusion is indeed subject-dependent, some subjects being more "auditory," and others more "visual." Intersubject variability has important consequences for theoretical understanding of the fusion process, and re-education of hearing impaired people. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3293001]

PACS number(s): 43.71.An, 43.71.Rt [KWG] Pages: 1584–1594

I. INTRODUCTION

When a public demonstration of the McGurk effect (McGurk and MacDonald, 1976) is presented to visitors or students, there appears a large variability in the subjects' audiovisual (AV) responses, some seeming focused on the auditory (A) input, others more sensitive to the visual (V) component and to the McGurk illusion. The existence of possible differences in fusion would have important consequences in both theoretical and practical terms. However, it stays hotly debated, considering that subjects could actually differ in pure auditory and visual performance rather than in fusion per se. In the following, the major elements of discussion and disagreement will be reviewed. Then it will be suggested that the debate has been largely blurred by methodological problems. A way out of these problems will be proposed, which will constitute the core of the present paper. The objective is actually twofold: present a methodological framework for analysis of audiovisual speech perception data and show that this framework confirms that there are indeed interindividual differences in the fusion process.

A. Interindividual differences in audiovisual fusion in the speech perception literature

The possibility that subjects could put more or less "weight" on the auditory or visual inputs is certainly not new. It was, for example, the focus of a paper by Seewald *et al.* (1985), suggesting that there was a "primary modality for speech perception," either auditory or visual. This hypothesis has received less attention since the work by Massaro and colleagues in the framework of the development of the "fuzzy-logical model of perception" (FLMP). Indeed, a central assumption of the model is that, apart from possible differences in auditory or visual perception, the fusion mechanism per se is exactly the same for all subjects (Massaro, 1987, 1998). The mechanism, actually a multiplicative process applied to fuzzy-logical levels of confidence provided by audition and vision on all possible answers, is considered as being an optimal process, in the sense that "all sources contribute to a decision but more ambiguous sources are given less of a say in the decision" (Massaro, 1998, p. 115).

A repeated claim by Massaro and colleagues is hence that all subjects are "optimal integrators" and combine auditory and visual evidence for available categories all exactly in the same multiplicative way. Any difference in the output of the audiovisual speech perception process would be only due to differences in auditory and visual processing and unisensory category tuning between subjects.

The hypothesis of a universal and optimal fusion mechanism remains controversial, and was the object of a series of experimental and modeling work by Grant and Seitz (1998), who claimed that, even when unimodal skill levels are taken into account, large differences in individuals' AV recognition scores persist which "might be attributable to differing efficiency in the operation of a perceptual process that *integrates* auditory and visual speech information" (p. 2438). The debate further continued between Massaro and Cohen (2000) and Grant (2002). Actually, the question of possible intersubject differences in AV integration, apart from being theoreti-

a)Author to whom correspondence should be addressed. Electronic mail: jean-luc.schwartz@gipsa-lab.inpg.fr

cally challenging, has important potential practical applications. Indeed, if some subjects integrate the audio and visual information less efficiently than others, the focus in a rehabilitation process (in case of hearing impairment for example) should be put on the training of integration, rather than just the training of auditory or visual abilities (Grant and Seitz, 1998). Incidentally, Rouger *et al.* (2007) claim to have found that cochlear implanted subjects are better at integrating the sound and face of a speaker's utterances than normal hearing subjects.

In the last 15 years, a number of studies have shown substantial individual variability in AV speech recognition. Sekiyama and Tokhura (1991) showed that McGurk fusion illusions were reduced in Japanese compared with English participants. Since then, several studies have investigated comparative language effects for audiovisual speech integration: English vs Japanese (Sekiyama and Tokhura, 1993; Kuhl *et al.*, 1994), English vs Japanese vs Spanish (Massaro *et al.*, 1993), Spanish vs German (Fuster-Duran, 1995), or German vs Hungarian (Grassegger, 1995). A number of differences have been reported. Some of them come from the nature of the stimuli, differing from language to language. However, there remain differences between linguistic groups perceiving the same stimuli. Sekiyama and Tokhura (1993) claimed that they reflect variations in the *weight* different linguistic communities would attribute to the visual input in the integration process. They suggested that the Japanese community could make less use of the visual input because of a cultural difference, namely, that "it may be regarded as impolite in Japan to look at someone's face" (p. 442). On the other hand, Massaro *et al.* (1993) and Kuhl *et al.* (1994) interpreted these differences as coming from variations in the inventory of linguistic prototypes rather than from social or cultural variations in the tuning of the audiovisual process. Indeed, Massaro *et al.* (1993) showed that their own data displaying different audiovisual perception of conflicting AV stimuli by English, Spanish, and Japanese subjects, could be perfectly fitted by FLMP. In the FLMP fit, the differences between English, Spanish, and Japanese subjects cannot be due to differences in fusion: they are totally due to differences in the unimodal categorization responses.

More recently, Sekiyama *et al.* (2003) showed that the very early ability to fuse auditory and visual inputs, displayed by a McGurk effect appearing as soon as 4 months in infants' speech perception (Burnham and Dodd, 1996, 2004; Rosenblum *et al.*, 1997), was followed by a developmental evolution of AV fusion after 6 years, and largely between 6 and 8 (Sekiyama and Burnham, 2004) for English children. This increase could be the result of a learning process, and it seems to be blocked in Japanese children, hence resulting in the smaller role of the visual input in AV perception previously described. Once again, however, it could be argued that this developmental pattern is just indicative of a development of *unimodal* auditory and visual categories rather than of integration per se. Basically, children (and particularly English ones) would be progressively more and more accurate in their perception of visual categories, hence the increase in AV performance. In this reasoning, fusion would stay perfectly stable whatever the age, i.e., multiplicative and optimal, in Massaro's sense.

Finally, gender differences in audiovisual fusion have been suggested in various papers, female subjects presenting a higher level of audiovisual performance and a greater level of visual influence on auditory speech (Irwin *et al.*, 2006; Strelnikov *et al.*, 2009), possibly linked to differences in the cortical networks involved, with less left lateralization in females compared with males (Pugh *et al.*, 1996; Jaeger *et al.*, 1998).

B. A methodological caveat: the 0/0 problem in FLMP testing

In a recent paper, Schwartz (2006) displayed a severe technical problem in the comparison of FLMP with other models when using corpora containing McGurk data. Indeed, in the case of conflicting inputs, the audio and visual stimuli provide at least one quasinull probability in each possible category, and the multiplicative process implied by the FLMP leads to AV predictions equal to 0/0, which is indeterminate. Therefore, any audiovisual response can be fitted by the FLMP. The consequence is double. First, since FLMP may predict any pattern of response in the McGurk case, fitting McGurk data with FLMP cannot help determine if variation in AV perception is actually due to differences in unimodal behavior or in AV fusion. Second, the overfitting ability of the FLMP with discrepant A and V stimuli might well contaminate the global root mean squared error (RMSE) criterion systematically used when FLMP is compared with other models. For these reasons, it seems more appropriate to use a Bayesian model selection (BMS) criterion, which intrinsically accounts for any overfitting problem (MacKay, 1992; Pitt and Myung, 2002; Schwartz, 2006).

In this context, the present paper aims at reconsidering the invariant vs subject-dependent audiovisual fusion problem, in a sound BMS framework. A classical test corpus of audiovisual consonant-vowel stimuli extensively studied by Massaro (1998) will provide a basis for assessing possible discrepancies in audiovisual fusion between subjects, independent of any linguistic or developmental effect. For this aim, weighted FLMP (WFLMP), a variant of FLMP explicitly incorporating subject-dependent weights of the audio and visual inputs in integration, will be compared with FLMP. This will provide the opportunity to use both BMS and RMSE criteria on these two models. This will also lead to a principled methodology for comparing audiovisual speech perception models on a given set of data. This methodology uses a so-called Laplace approximation of BMS, called BMSL, together with a systematic assessment of the number of really useful parameters in the models to compare, relying on the BMS ability to deal with variations in the number of degrees of freedom in these models.

Section II will recall the experimental material and provide a detailed description of the proposed methodology, together with the models to compare, and the assessment criteria. In Sec. III, the obtained results will be presented, before a discussion in Sec. IV.

II. METHODOLOGY

A. Experimental material: The UCSC corpus of CV audiovisual discrepant stimuli

The corpus considered here has been extensively used for comparing audiovisual fusion models in speech perception (Massaro, 1998). This corpus crosses a synthetic five-level audio /ba/-/da/ continuum with a synthetic video similar continuum. The 10 unimodal (5A, 5V) and 25 bimodal (AV) stimuli were presented for /ba/ vs /da/ identification to 82 subjects, with 24 observations per subject. The responses are kindly made available by Massaro and colleagues on their website (http://mambo.ucsc.edu/ps1/8236/).

B. Model comparison

1. RMSE and corrected RMSE

Let us consider a given speech perception experiment consisting in the categorization of speech stimuli involving n_E experimental conditions Ej, and in each condition, n_C possible responses corresponding to different phonetic categories C_i. In most papers comparing models in the field of speech perception, the tool used to compare models is the "fit" estimated by the RMSE, computed by taking the squared distances between observed and predicted probabilities of responses, averaging them over all categories C_i and all experimental conditions Ej, and taking the square root of the result

$$\mathrm{RMSE} = \left[\left(\sum_{\mathrm{Ej},C_i} (P_{\mathrm{Ej}}(C_i) - p_{\mathrm{Ej}}(C_i))^2\right)/(n_E n_C)\right]^{1/2} \quad (1)$$

(observed probabilities are in lower case and predicted probabilities in upper case throughout this paper).

Considering that two models M_A and M_B might differ in their number of degrees of freedom, Massaro (1998) proposed to apply a correction factor $k/(k-f)$ to RMSE, with k the number of data and f the number of degrees of freedom of the model (p. 301). This provides a second criterion:

$$\mathrm{RMSEcor} = k/(k-f)\left[\left(\sum_{\mathrm{Ej},C_i} (P_{\mathrm{Ej}}(C_i) - p_{\mathrm{Ej}}(C_i))^2\right)/(n_E n_C)\right]^{1/2}. \quad (2)$$

2. BMSL

If $\mathbf{D}$ is a set of k data d_i, and M a model with parameters $\boldsymbol{\Theta}$, the fit may be derived from the logarithm of the *maximum likelihood of the model* considering the data set, that is the value of $\boldsymbol{\Theta}$ maximizing $L(\boldsymbol{\Theta}|M) = p(\mathbf{D}|\boldsymbol{\Theta},M)$. However, comparing two models by comparing their best fits means that there is a first step of estimation of these best fits, and it must be acknowledged that the estimation process is not error-free. Therefore, the comparison must account for this error-prone process, which is done in Bayesian model selection by computing the total likelihood of the model knowing the data. This results in integrating likelihood over all model parameter values. Taking the opposite of the logarithm of total likelihood leads to the so-called BMS criterion that should be minimized for model evaluation (MacKay, 1992; Pitt and Myung, 2002):[1]

$$\mathrm{BMS} = -\log \int L(\boldsymbol{\Theta}|M)p(\boldsymbol{\Theta}|M)d\boldsymbol{\Theta}. \quad (3)$$

The computation of BMS through Eq. (3) is complex. It involves the estimation of an integral, which generally requires use of numerical integration techniques, typically Monte Carlo methods (e.g., Gilks *et al.*, 1996). However, Jaynes (1995) (Chap. 24) proposed an approximation of the total likelihood in Eq. (9), based on an expansion of log(L) around the maximum likelihood point $\boldsymbol{\theta}$:

$$\log(L(\boldsymbol{\Theta})) \cong \log(L(\boldsymbol{\theta})) + 1/2(\boldsymbol{\Theta} - \boldsymbol{\theta})'[\partial^2 \log(L)/\partial\boldsymbol{\Theta}^2]_{\boldsymbol{\theta}}(\boldsymbol{\Theta} - \boldsymbol{\theta}), \quad (4)$$

where $[\partial^2 \log(L)/\partial\boldsymbol{\Theta}^2]_{\boldsymbol{\theta}}$ is the Hessian matrix of the function log(L) computed at the position of the parameter set $\boldsymbol{\theta}$ providing the maximal likelihood L_{max} of the considered model. This leads to the so-called Laplace approximation of the BMS criterion (Kass and Raftery, 1995):

$$\mathrm{BMSL} = -\log(L_{max}) - m/2\log(2\pi) + \log(V) - 1/2\log(\det(\boldsymbol{\Sigma})), \quad (5)$$

where V is the total volume of the space occupied by parameters $\boldsymbol{\Theta}$, m is its dimension, which is the number of free parameters in the considered model, and $\boldsymbol{\Sigma}$ is defined by

$$\boldsymbol{\Sigma}^{-1} = -[\partial^2 \log(L)/\partial\boldsymbol{\Theta}^2]_{\boldsymbol{\theta}}. \quad (6)$$

The preferred model considering the data $\mathbf{D}$ should *minimize* the BMSL criterion. There are, in fact, three kinds of terms in Eq. (5). First, the term $-\log(L_{max})$ is directly linked to the maximum likelihood of the model, more or less accurately estimated by RMSE in Eq. (1): the larger the maximum likelihood, the smaller the BMSL criterion. Then, the two following terms are linked to the dimensionality and volume of the considered model. Altogether, they result in handicapping models that are too "large" (that is, models with a too high number of free parameters) by increasing BMSL.[2] Finally, the fourth term provides a term favoring models with a large value of $\det(\boldsymbol{\Sigma})$. Indeed, if $\det(\boldsymbol{\Sigma})$ is large, the determinant of the Hessian matrix of log(L) is small, which expresses the fact that the likelihood L does not vary too quickly around its maximum value L_{max}. This means that the fit provided by the model around its maximum likelihood point is stable: exactly the contrary of FLMP with McGurk data, since its overfitting abilities result in very rapid modifications of the prediction even for very small changes in the unimodal values, making the integration process quite unstable and oversensitive to the tuning of free parameters in the model. Derivation of the exact formula in Eq. (5), together with a practical implementation of BMSL, can be found in http://www.icp.inpg.fr/~schwartz/fichiers_pdf/BMSL_tutorial.pdf.

Bayesian model selection has already been applied to the comparison of AV speech perception models, including FLMP (see Myung and Pitt, 1997; Massaro *et al.*, 2001; Pitt *et al.*, 2003). However, this involved heavy computations of integrals in Eq. (3) through Monte Carlo techniques, which would be difficult to apply systematically in model compari-

sons. BMSL has the advantage of being easy to compute and to interpret in terms of fit and stability. Furthermore, if the amount of available data is much higher than the number of parameters involved in the models to compare (that is, the dimension m of the Θ space), the probability distributions become highly peaked around their maxima, and the central limit theorem shows that the approximation in Eqs. (4) and (5) becomes quite reasonable (Walker, 1967). Kass and Raftery (1995) suggested that the approximation should work well for a sample size greater than 20 times the parameter size m (see Slate, 1999, for further discussions about assessing non-normality).

3. Estimating the "true" number of degrees of freedom in a model

The number of model parameters in most model comparison studies in AV speech perception is generally kept fixed to the "natural number of degrees of freedom" of the model, that is, the number of free parameters necessary to implement the model in its most extensive definition. Care is generally taken to check that the models have basically the same number of degrees of freedom; otherwise the RMSE correction described previously could be applied. Notice that this correction loses some sense if a parameter is introduced with no effect on the model likelihood (a "useless parameter") while BMSL naturally discards useless parameters through the integration in formula (3).

Of course, completely useless parameters generally do not exist, since this would correspond to some kind of misconception of the model. However, it is important to assess the possibility that some parameters are not really useful in the model behavior. For example, while all model comparisons generally involve a subject-by-subject assessment—and it will also be the case here—it could be interesting to test if some parameters could not, in fact, be similar from one subject to the other. The same could be done from one experimental condition to the other. Therefore, various implementations of the models to compare will be systematically tested, with a progressively increasing number of fixed parameters and thus a decreasing number of free parameters, in order to attempt to determine the *true* number of degrees of freedom of the model, that is, the number of free parameters really useful, and providing the highest global likelihood of the model knowing the data. Our basic assumption is that it is under the condition of true number of degree of freedom that models can be really assessed and compared in sound conditions.

Decreasing the number of free parameters raises two problems. First, the parameters to fix must be adequately selected. This may be done on a statistical objective basis, for example, through principal component analysis techniques, but this results in combinations of parameters difficult to interpret. A heuristic approach was preferred in which the observation of experimental data guided the selection of possible parameters to be kept fixed from one subject to another. The second problem is to estimate the value of the parameters being kept fixed. This was done through a Round Robin technique, in which a given parameter for one subject is estimated from the mean value taken by the parameter in the whole corpus excluding the current subject from the computation. This technique, classical and computationally simple, prevents from any artifactual introduction of the current data to model inside the "fixed" parameter used to model the data in a circular approach, which would be inappropriate.

C. Models

Two models were compared, FLMP and a variant with weighted contribution of the auditory and visual inputs in the integration, WFLMP. For each corpus, each model (including the variants associated with the decrease in the number of degrees of freedom) was fitted to the data separately for each subject. This enabled us to compute both mean values of the selected criteria, averaged over all subjects, and to assess differences between models by applying Wilcoxon signed-rank tests over the compared criteria for each subject.

1. FLMP

In a speech perception task consisting in the categorization of auditory, visual, and audiovisual stimuli, the FLMP may be defined as a Bayesian fusion model with independence between modalities, and the basic FLMP equation is

$$P_{AV}(C_i) = P_A(C_i)P_V(C_i)/\sum_j P_A(C_j)P_V(C_j), \quad (7)$$

C_i and C_j being phonetic categories involved in the experiment, and P_A, P_V, and P_{AV} the model probability of responses respectively in the A, V, and AV conditions.

2. WFLMP

The weighted FLMP model, called WFLMP, is defined by

$$P_{AV}(C_i) = P_A^{\lambda_A}(C_i)P_V^{\lambda_V}(C_i)/\sum_j P_A^{\lambda_A}(C_j)P_V^{\lambda_V}(C_j), \quad (8)$$

where λ_A and λ_V are subject-dependent factors used to weight the A and V inputs in the computation of the audiovisual responses estimated by $P_{AV}(C_i)$ (see other introductions of weights inside FLMP in Schwarzer and Massaro, 2001; or for a similar kind of weighted fusion model applied to speech recognition, in various implementations since Adjoudani and Benoît, 1996: see a review in Teissier *et al.*, 1999). For each subject, a lambda value is defined between 0 and 1, and λ_A and λ_V are computed from lambda by: $\lambda_A = \text{lambda}/(1-\text{lambda})$ and $\lambda_V = (1-\text{lambda})/\text{lambda}$, with thresholds maintaining λ_A and λ_V between 0 and 1. Figure 1 shows how lambda controls the weights λ_A and λ_V and how this results in varying P_{AV} from a value close to P_A when lambda is close to 0, to a value close to P_V when lambda is close to 1, passing by a value identical to the FLMP prediction when lambda is set at 0.5, with λ_A and λ_V both equal to 1.

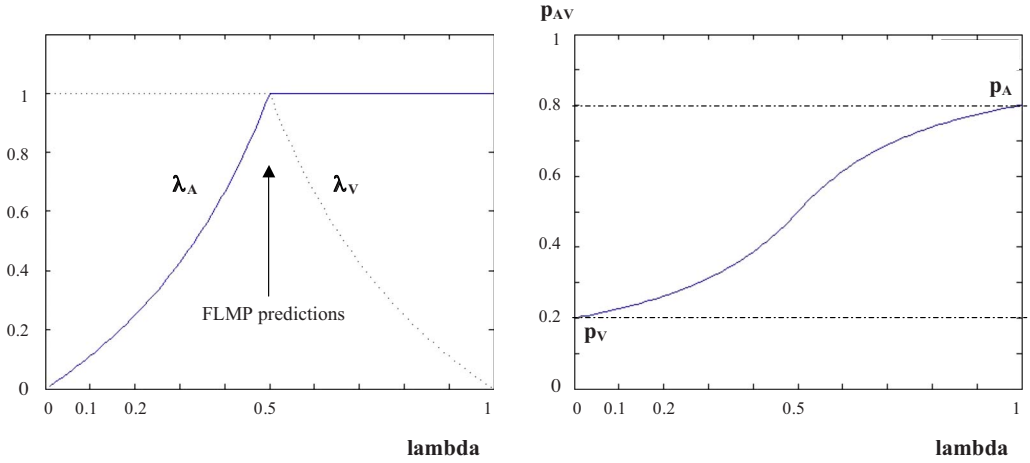


FIG. 1. (Color online) Variations in weighting coefficients λ_A and λ_V (left) and predicted p_{AV} (right) as a function of the lambda parameter tuning fusion in WFLMP. When lambda is close to 0, the audio weight decreases toward zero, the video weight increases toward 1, and the modeled p_{AV} reaches a value close to p_V. Conversely, when lambda is close to 1, the audio weight increases toward 1, the video weight decreases toward zero, and the modeled p_{AV} reaches a value close to p_A. Notice that for a lambda value at 0.5, both audio and video weights are set to 1, which provides exactly the FLMP predictions. In this example, p_V is set to 0.2 and p_A to 0.8.

III. RESULTS

A. Analysis of individual experimental data

The UCSC corpus has been extensively used in AV speech perception model assessment, generally with a good fit using the FLMP and RMSE criterion (Massaro, 1998; see also Massaro *et al.*, 2001, for an assessment of FLMP with a BMS criterion on this corpus). However, looking at the data, there seems to appear an effect not predicted by the FLMP, that is, interindividual differences in AV interaction. This is displayed in Fig. 2, showing two subjects with very close auditory and visual performances, although with quite different audiovisual responses. It seems that the weight of the visual modality is, respectively, high for the first one [Fig. 2(a)] and low for the second one [Fig. 2(b)]. Though the FLMP does not incorporate A and V weights, the fit is, however, quite acceptable (with RMSE values, respectively, 0.04 and 0.02 for these two subjects). This good fit is actually obtained because of the 0/0 instability: indeed, the FLMP simulation of unimodal data for the first subject is drawn toward slightly more ambiguous values for A responses and less ambiguous values for V responses [see Fig. 2(a)], while the inverse is done for the second subject [see Fig. 2(b)]. This is the indirect way the FLMP may decrease the importance of a modality in fusion, by slightly but consistently misfitting the unimodal data without introducing subject-specific weights, and while keeping a very low RMSE value (a very good fit) because of the 0/0 problem (Schwartz, 2006). Such consistent misfits of unimodal data, if they happen in a significant number of cases, would indicate a problem in modeling. They should be taken into account in a BMS criterion, although they are almost undetectable in a RMSE criterion.

B. Selected degrees of freedom for FLMP and WFLMP

The first implementation of FLMP needs ten parameters for each subject, that is five values $Ai=P_{Ai}(/da/)$ and five values $Vj=P_{Vj}$ (/da/) for the five stimuli of each continuum. Since the WFLMP model needs one more parameter per sub-

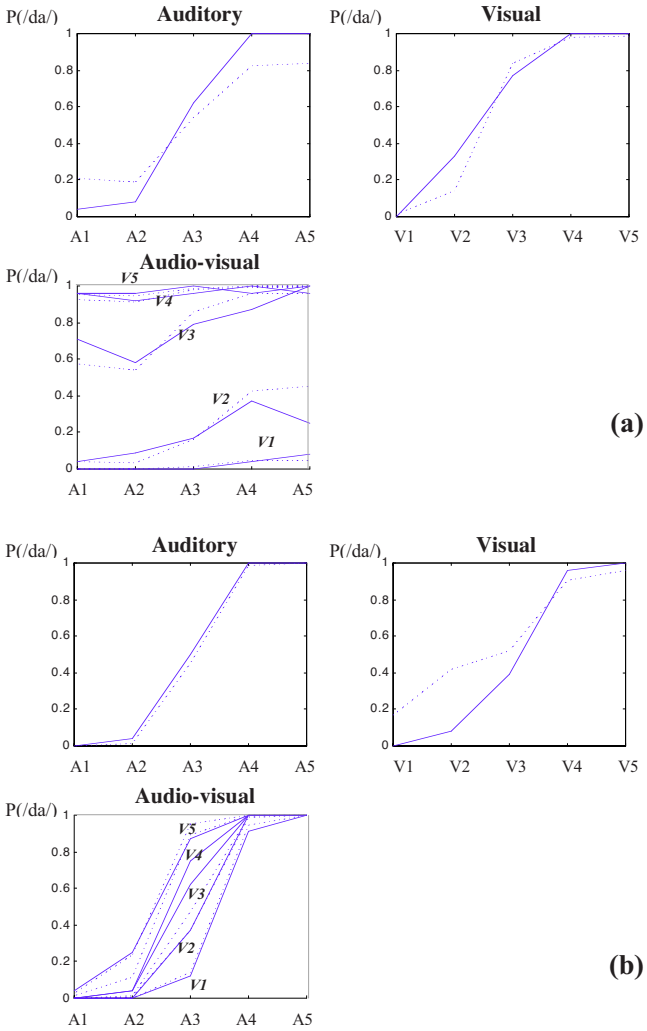


FIG. 2. (Color online) (a) Audio (top left), visual (top right), and audiovisual (bottom) data for subject 3 in UCSC corpus: data in solid lines and FLMP predictions in dotted lines. (b) Same as Fig. 2(a) for subject 18 in UCSC corpus.

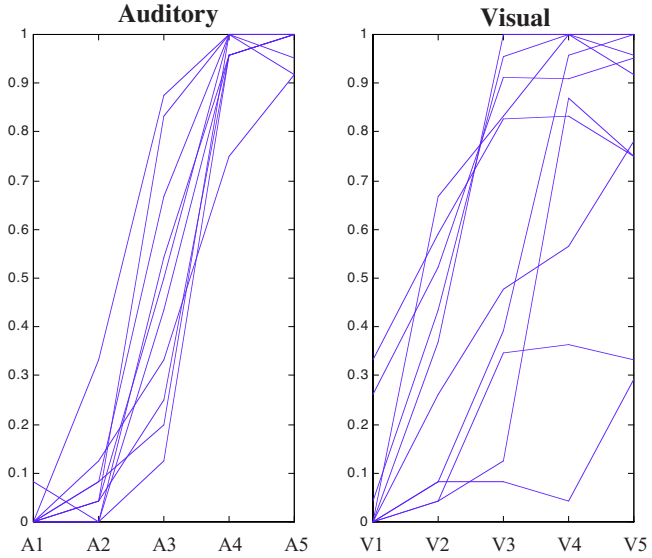


FIG. 3. (Color online) Left: audio identification for ten subjects. Right: visual identification for ten subjects.

ject, one parameter was removed by fixing the value of the parameter A5 (audio response for the fifth audio stimuli, higher than 0.99 in average) at a value equal to the mean of the value it takes for the other subjects.

To explain how the number of parameters was decreased, on Fig. 3 a sample of auditory and visual identification curves is displayed for 10 of the 82 subjects. In the audio results [Fig. 3(a)], the curves are all S-shaped from a value close to 0 to a value close to 1, with less variation on the sides (for A5, A1, and to a lesser extent A4 and A2). Therefore, it was attempted to fix these parameters, in this order, with the Round Robin procedure. In the visual curves [Fig. 3(b)], the configuration is different and suggests that it should be possible to describe these curves by estimating some values by a linear regression prediction on logit values of Vi, that is, $\log(Vi/(1-Vi))$. For this aim, two linear regression predictions on logit values were defined, one predicting V2 and V4 from the parameters V1, V3, and V5, and the other predicting V2, V4, and V5 from the parameters V1 and V3. Altogether, this lead to five variants of the FLMP and WFLMP models, respectively, with 10, 6, 5, 4, and 3 free parameters per subject (Table I).[3]

C. Modeling results

Figure 4 shows the results for the two models with their five free-parameter variants. For each case, means and standard deviations computed on the modeling results for the 82 subjects are presented.

With ten parameters per subject, the FLMP fit is good, with an average RMSE value at 0.051 (see Massaro, 1998, p. 64). Interestingly, the fit is significantly better for the WFLMP with the same number of free parameters, with an average RMSE value at 0.0445 (since $N=82$ is higher than 20, z-ratios are used following a unit normal distribution, $u=4.92$, $p<0.001$). RMSE then logically increases when the number of parameters decreases [Fig. 4(a)]. The portrait for RMSEcor is the same [Fig. 4(b)]. However, BMSL reaches a minimum for six parameters, both for FLMP and WFLMP [Fig. 4(c)]. In this variant of both models, A1 and A5 are fixed, together with A4 for WFLMP. V2 and V4 are estimated from V1, V3, and V5 by logit linear regression. For this optimal six-parameter implementation, there is a significant gain of WFLMP over FLMP ($u=4.77$, $p<0.001$).

TABLE I. The five variants of FLMP and WFLMP. All fixed parameters are estimated by the Round Robin technique.

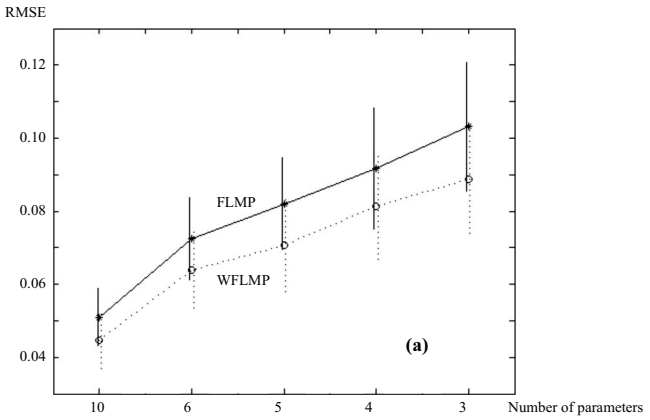
No. of parameters per subject	Parameters for FLMP	Parameters for WFLMP
10	V1–V5, A1–A5	+lambda A5 fixed
6	V1, V3, V5, A2, A3, A4 V2, V4 estimated by linear regression A1, A5 fixed	+lambda A4 fixed
5	V1, V3, A2, A3, A4 V2, V4, V5 estimated by linear regression A1, A5 fixed	+lambda A4 fixed
4	V1, V3, A2, A3 V2, V4, V5 estimated by linear regression A1, A4, A5 fixed	+lambda A2 fixed
3	V1, V3, A3 V2, V4, V5 estimated by linear regression A1, A2, A4, A5 fixed	+lambda V1 fixed

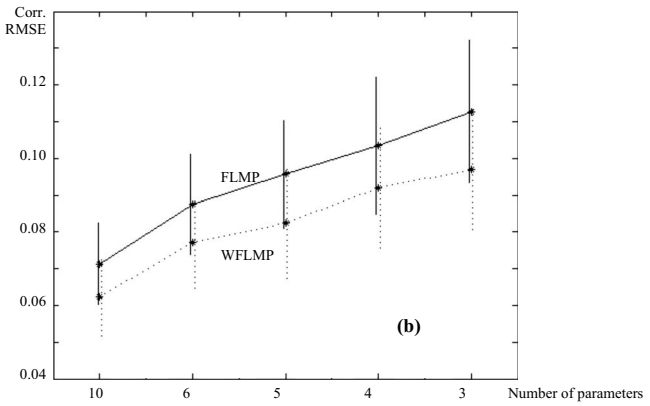
In Fig. 5, the histogram of logarithms of estimated lambda values are plotted for all subjects for WFLMP with six parameters. It appears that the range is indeed large, with auditory subjects on the right of the 1-value and visual subjects on the left. Under a criterion of λ_A/λ_V values, respectively, higher than 1.5 or lower than 0.67 (1/1.5), there are 33 "audio" and 14 "visual" subjects, the remaining 35 being intermediary.

Figure 6 shows how WFLMP models the two subjects compared in Fig. 2. Typically, the auditory and visual fits are similar between subjects—as in the experimental data themselves—while the good fit of the differences between subjects in audiovisual values is due to large differences in the lambda values, as shown on Fig. 5. This confirms that subject (a) is rather visual (with a λ_A/λ_V ratio at 0.23) and subject (b) is rather auditory (with a λ_A/λ_V ratio at 3.89), as suggested by the data themselves.

IV. GENERAL DISCUSSION

Two topics are addressed in the present work. One concerns methodology for model comparison, which is of particular importance in audiovisual fusion, as evidenced by the very large number of controversies regularly arising in the domain. This also has implications for designing models for audiovisual fusion in speech perception. The second one concerns the invariant vs subject-dependent nature of audiovisual fusion and more generally the parameters able to intervene in fusion. These topics will be addressed one after the other.





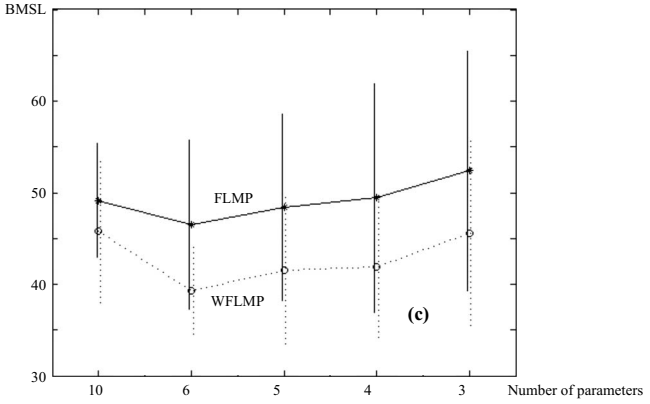


FIG. 4. (a) Compared RMSE values for FLMP vs WFLMP simulations with three to ten free parameters per subject. (b) Compared corrected RMSE values for FLMP vs WFLMP simulations with three to ten free parameters per subject. (c) Compared BMSL values for FLMP vs WFLMP simulations with three to ten free parameters per subject.

A. An adequate methodology for comparing models

There are two important claims in our methodological approach. First, a local criterion such as RMSE, or its quasi-equivalent maximum likelihood criterion, can be inappropriate, particularly in cases involving models which have a tendency to overfit the data, e.g., with FLMP and McGurk data. A global BMS criterion is theoretically sounder, as has been discussed in a large number of papers, unfortunately not much acknowledged in the speech perception community. The local approximation provided by BMSL is simple to compute, easy to interpret, and efficient assuming that the number of experimental data points is sufficient. In the present paper, it appears that RMSE and BMSL converge on showing the superiority of WFLMP over FLMP for the McGurk data. But this is not systematically the case (see,

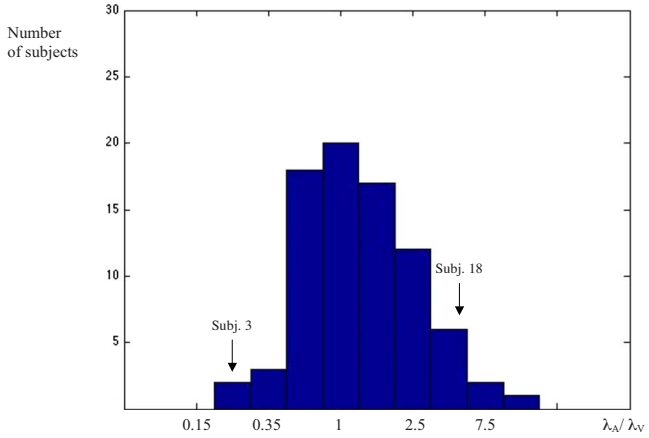


FIG. 5. (Color online) Histogram of λ_A/λ_V values controlling the fusion process for the 82 subjects in the WFLMP model with six parameters. Values for subject 3 [data displayed in Fig. 1(a)] and subject 18 [data displayed in Fig. 1(b)] are superimposed on the figure.

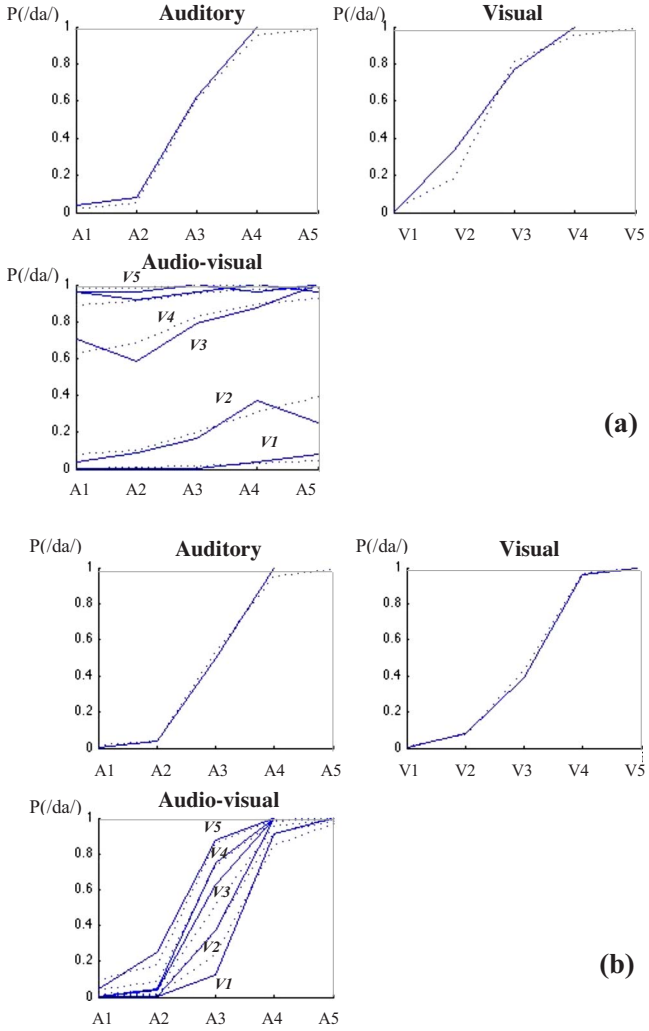


FIG. 6. (Color online) (a) Audio (top left), visual (top right), and audiovisual (bottom) data for subject 3 in UCSC corpus: data in solid lines and predictions with WFLMP with six parameters in dotted lines. (b) Same as Fig. 6(a) for subject 18 in UCSC corpus.

e.g., Schwartz and Cathiard, 2004; Schwartz, 2006). Therefore, we suggest that any model comparison involving FLMP based on a RMSE criterion should be taken cautiously, and its conclusions should be considered as probably arguable unless a new evaluation based on Bayesian model selection is undertaken.

Second, varying the number of parameters in model assessment is very important. Of course, the difficulty is that there is much freedom in the strategies that can be proposed for this principled reduction. This should involve both a simple and intuitive approach, and a substantial number of variants to be able to assess the approximate number of really meaningful parameters for comparing models.

In fact, these two claims are related. Decreasing the number of parameters *forces* one to use a criterion able to take the size of the model parameter set into account. This is the case of BMS and its BMSL approximation, and *not* the case of RMSE, in which any correction is largely arbitrary. Variation in the number of parameters showed that there was indeed some redundancy in the ten free parameters per subject involved in FLMP for the present corpus, six appearing as a more plausible number of degrees of freedom (three for the visual input, three for the audio input). It is interesting to note that a reduced number of degrees of freedom provided a much larger difference in favor of WFLMP, compared with the complete set of ten parameters.

B. Audiovisual fusion models

The present paper is focused on FLMP for methodological reasons associated with its very frequent use in publications, and its excessive adaptability to McGurk data leading in several cases to inappropriate or mistaken analyses of experimental results. However, FLMP is actually neither weakened nor strengthened by the present paper.

It is not weakened since we proposed both an adequate method for testing it in safer conditions—through the BMS framework—and a possible variant with subject-specific modulation—WFLMP—which could provide the route for new developments more in line with evidence that fusion is subject-dependent. Massaro and colleagues already introduced both ingredients in some of their work, but the present paper shows that they are actually *required* in any further use of FLMP in speech perception, particularly (but not exclusively) in experiments involving incongruent stimuli, as in the McGurk effect.

It is not strengthened either, since the present analysis could have been applied to other fusion models, such as Braida's pre- and postlabeling models (Braida, 1991; see also Grant and Seitz, 1998; Grant, 2002), with quite probably the same conclusions. Actually, a number of models have been recently developed explicitly taking into account the possibility that one modality could be favored in the fusion process. This is the point addressed by Ernst and colleagues with their maximum likelihood framework according to which integration would be "optimal" in leading to the largest possible reduction in the variance of the multisensory output. This is achieved by adaptively weighting modalities in relation to their reliability (or variance) for the considered task (Ernst and Banks, 2002; Ernst and Bulthoff, 2004; Denève and Pouget, 2004; Körding *et al.*, 2007 and a precursor use of this concept in audiovisual speech perception models in Robert-Ribes *et al.*, 1995). This "optimal integration" view is different from the optimal Bayesian fusion of decisions implemented in FLMP, since it occurs at a precategorical level.

C. What drives audiovisual fusion in speech perception?

The major theoretical output of the present paper is that it clearly shows that fusion is subject-dependent. There are indeed large intersubject differences in audiovisual fusion for speech perception, with various groups of subjects, some being more auditory, and others more visual. Many papers mention such a large variability in audiovisual performance. However, it was always unclear whether this was due to differences in unisensory performance or multisensory integration. The present analysis strongly reinforces the second view.

This opens the route to a number of questions about the fusion mechanism itself. Differences between subjects in the McGurk paradigm could result from a general "orientation" of a given subject toward one or the other modality for individual reasons (specific or related to, e.g., culture, language, sex, or age). They could also be the consequence of properties of the task or the experimental situation, which could have driven the subject toward one rather than the other stimulus input in a bimodal task.

1. Interindividual factors

It could well be the case that some subjects rely more on audition and others more on vision, and that they weight audiovisual fusion accordingly (see Giard and Peronnet, 1999). Hence, it could be assumed that there is for a given subject a general trend to favor one modality over another one, whatever the task. This should result in future studies comparing audiovisual fusion in various speech and nonspeech tasks, searching for individual portraits stable from one task to the other. These different behaviors could also be associated with differences in neuroimagery experiments, in terms of the involved cortical networks and of the quantitative role of each part in the global portrait.

We have already discussed in the Introduction possible factors that likely to play a role in sensor fusion: some languages could use the visual input more than others (e.g., English more than Japanese), female subjects could use it more than males, and adults more than children. Notice that the methodology employed here could be used to reanalyze all data relevant for these claims, in order to carefully disentangle the role of unimodal and multimodal factors in the corresponding studies.

This opens an important question, which is to know to what extent the weighting process can be dynamically modified during the subject's life. We have already mentioned the developmental evolution leading to an increase in the role of the visual input (see e.g., Sekiyama *et al.*, 2003; Sekiyama and Burnham, 2004). Recent data by Schorr *et al.* (2005)

suggested that there is a critical period for the development of audiovisual fusion in speech perception, before 2.5 years. In the case of a perturbation of one or the other modality, related to age or handicap, the question becomes to know if a subject can, voluntarily or through any reeducation means, selectively reinforce the weight of the most efficient modality.

In a recent study, Rouger *et al.* (2007) claimed that this could indeed be the case, hearing impaired subjects equipped with cochlear implants displaying, in their terms, "a greater capacity to integrate visual input with the distorted speech signal" (p. 7295). Actually, these data should be considered with caution, being a possible case of unimodal effects interpreted as bimodal. Indeed, Rouger *et al.* compared three populations of subjects: hearing impaired subjects equipped with cochlear implants (CIs), normal hearing subjects with audition degraded by noise (NHN), and normal hearing subjects presented with noise-band vocoder speech degrading audition in a way supposed to mimic the cochlear implant (NHV). They showed that for a similar level of audio performance, the audiovisual recognition is larger for CI than for NHV, NHN being in the middle. Two factors could be, in their view, responsible for this pattern: differences in the visual performance and in integration per se. A modeling approach leads them to claim that while the global visual scores are actually better in CI compared with NHN and NHV, there would be an additional gain in CH compared with NHV, hence the claim about a "greater capacity to integrate visual input with the distorted speech signal." Notice that integration efficiency would be as high in NHN as in CH according to their analysis. However, careful inspection of auditory confusion matrices for NHV and NHN, available in Rouger, 2007, shows that the structure of these matrices is quite different. Importantly, the transmission of the voicing mode was poorer in speech degraded with noise-band vocoder (NHV) than with white noise (NHN), suggesting that there could be a poorer complementarity in the audio and visual inputs in NHV, logically resulting in lower audiovisual scores. Differences in audiovisual performance would hence result from the structure of the unimodal inputs (being less complementary for normal hearing subjects presented with noise-band vocoder speech) rather than from integration per se. Actually, in this case, a study based on WFLMP and BMS would probably *not* reveal any discrepancy in integration between impaired subjects equipped with cochlear implants (CI) and normal hearing subjects (NHV and NHN).

2. Intraindividual factors

Finally, it is quite possible that the weight of one modality depends on the experimental situation per se. First, stimuli themselves could possibly drive the weighting factor. In a review of intersensory interactions, Welch and Warren (1986) proposed "modality precision" or "modality appropriateness" as a key factor in explaining which modality should dominate intersensory judgments. Evidence for the role of reliability in audiovisual fusion for speech perception can be found in the study by Lisker and Rossi, 1992 on the auditory, visual, and audiovisual identification of vocalic rounding. Careful inspection of their data shows that although auditory identification seems in some cases quite accurate, there is a systematic trend for putting more weight in the visual modality within audiovisual data, as if the subjects "knew" that their eye was better than their ear at this particular task. Conversely, Robert-Ribes *et al.* (1998) on their study of audiovisual vowel perception in noise reported that with a very high level of noise, some subjects consistently select a given response (e.g., [ø]) for all vowels in noise in the auditory modality, which could lead in a model as FLMP to a large probability of response of this category in the audiovisual modality. However, this is not the case, showing that subjects know that the auditory modality is not reliable at high levels of noise and hence discards it almost completely from the fusion process.

Second, the experimental conditions could lead to enhance or decrease the role of one modality in the fusion process. Attentional mechanisms should play a role at this level. Actually, while it had been initially claimed since McGurk and MacDonald, 1976 that the McGurk effect was automatic and not under the control of attention, it appeared later that the instruction to attend more to audition or to vision might bias the response (Massaro, 1998). A recent set of experiments by Tiippana *et al.* (2004) showed that if the attention is distracted from the visual flow, the role of the visual input seems to decrease in fusion, with less McGurk effect. Notice that the authors themselves attempted to simulate their data with FLMP and argued that the good fit of their data by a model claiming that fusion is automatic appeared as "a paradox" (p. 458). Actually, reanalysis of their data in a BMS framework with a weighted FLMP suggests that there are indeed attentional factors intervening in fusion itself, independent of unimodal effects (Schwartz and Tiippana, in preparation). This is confirmed by a number of recent experiments showing the possibility to modulate the McGurk effect based on manipulations of attention (e.g., Alsius *et al.*, 2005, 2007), although here again, a precise analysis of experimental results in a BMS framework could provide an adequate control for disentangling unimodal from multimodal factors.

V. CONCLUSION

The present work proposed a new methodology for comparatively assessing models of audiovisual speech perception. This methodology is based on both the use of a Bayesian model selection criterion approximated in a computationally simple way (BMSL) and on a systematic variation in the number of degrees of freedom of the models to assess, in order to reveal the "true" number of parameters in a given model for a given task. The comparison of FLMP with a variant with auditory and visual weights varying from one subject to another (WFLMP) leads to the conclusion that weights are indeed variable, and hence that audiovisual integration seems subject-dependent.

This could have important consequences in future studies about audiovisual speech perception. First, from a methodological point of view, it suggests that studies on audiovisual speech perception should consider these differences and possibly separate experimental groups into "auditory" or "vi-

sual" subgroups based on such criteria such as McGurk performance. Second, from an audiological point of view, this indicates that subjects should be assessed based on their audiovisual fusion abilities, and considered differently—in terms of reeducation and practice—depending on whether they are more auditory or more visual in their behavior.

[1]In the following, bold symbols deal with vectors or matrices, and all optimizations are computed on the model parameter set $\boldsymbol{\Theta}$.

[2]The interpretation of the term log(V) is straightforward and results in handicapping large models by increasing BMSL. The term $m/2\log(\pi)$ comes more indirectly from the analysis and could seem to *favor large models*. In fact, it can only decrease the trend to favor small models over large ones.

[3]It could seem paradoxical to maintain the number of free parameters similar for each subject, while attempting to show interindividual differences. This is not the case actually. The principle is to freeze as much as possible the structure of the model for all subjects, in order to let the differences appear in an objective way.

Adjoudani, A., and Benoît, C. (**1996**). "On the integration of auditory and visual parameters in an HMM-based ASR," in *Speechreading by Humans and Machines*, edited by D. G. Stork and M. E. Hennecke (Springer-Verlag, Berlin), pp. 461–472.

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (**2005**). "Audiovisual integration of speech falters under high attention demands," Curr. Biol. **15**, 839–43.

Alsius, A., Navarra, J., and Soto-Faraco, S. (**2007**). "Attention to touch weakens audiovisual speech integration," Exp. Brain Res. **183**, 399–404.

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. **43**, 647–677.

Burnham, D., and Dodd, B. (**1996**). "Auditory-visual speech perception as a direct process: The McGurk effect in human infants and across languages," in *Speechreading by Humans and Machines*, edited byD. G. Stork and M. E. Hennecke (Springer-Verlag, Berlin), pp. 103–114.

Burnham, D., and Dodd, B. (**2004**). "Auditory-visual speech integration by pre-linguistic infants: Perception of an emergent consonant in the McGurk effect," Dev. Psychobiol. **44**, 209–220.

Denève, S., and Pouget, A. (**2004**). "Bayesian multisensory integration and cross-modal spatial links," J. Neurophysiol. **98**, 249–258.

Ernst, M. O., and Banks, M. S. (**2002**). "Humans integrate visual and haptic information in a statistically optimal fashion," Nature (London) **415**, 429–433.

Ernst, M. O., and Bulthoff, H. H. (**2004**). "Merging the senses into a robust percept," Trends Cogn. Sci. **8**, 162–169.

Fuster-Duran, A. (**1995**). "McGurk effect in Spanish and German listeners. Influences of visual cues in the perception of Spanish and German conflicting audio-visual stimuli," in Proceedings of the Eurospeech 95, pp. 295–298.

Giard, M. H., and Peronnet, F. (**1999**). "Auditory-visual integration during multi-modal object recognition in humans: a behavioral and electrophysiological study," J. Cogn. Neurosci. **11**, 473–490.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (**1996**). *Markov Chain Monte Carlo in Practice* (Chapman and Hall, New York).

Grant, K. W. (**2002**). "Measures of auditory-visual integration for speech understanding: A theoretical perspective (L)," J. Acoust. Soc. Am. **112**, 30–33.

Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2449.

Grassegger, H. (**1995**). "McGurk effect in German and Hungarian listeners," in Proceedings of the 13th International Congress on Phonetic Sciences, pp. 210–113.

Irwin, J. R., Whalen, D. H., and Fowler, C. A. (**2006**). "A sex difference in visual influence on heard speech," Percept. Psychophys. **68**, 582–592.

Jaeger, J., Lockwood, A., Van Valin, R. D., Jr., Kemmerer, D. L., Murphy, B. W., and Wack, D. S. (**1998**). "Sex differences in brain regions activated by grammatical and reading tasks," NeuroReport **9**, 2803–2807.

Jaynes, E. T. (**1995**). *Probability Theory: The logic of Science* (Cambridge University Press, Cambridge).

Kass, R. E., and Raftery, A. E. (**1995**). "Bayes factor," J. Am. Stat. Assoc. **90**, 773–795.

Körding, K. P., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., and Shams, L. (**2007**). "Causal inference in cue combination," PLoS ONE **2**, e943.

Kuhl, P. K., Tsuzaki, M., Tohkura, Y., and Meltzoff, A. N. (**1994**). "Human processing of auditory-visual information in speech perception: Potential for multimodal human-machine interfaces," in Proceedings of International Conference on Spoken Language Processing, Yokohama, Japan, pp. 539–542.

Lisker, L., and Rossi, M. (**1992**). "Auditory and visual cueing of the [± rounded] feature of vowels," Lang Speech **35**, 391–417.

MacKay, D. J. C. (**1992**). "Bayesian interpolation," Neural Comput. **4**, 415–447.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Laurence Erlbaum Associates, London).

Massaro, D. W. (**1998**). *Perceiving Talking Faces* (MIT, Cambridge, Ma).

Massaro, D. W., and Cohen, M. M. (**2000**). "Tests of auditory-visual integration efficiency within the framework of the fuzzy-logical model of perception," J. Acoust. Soc. Am. **108**, 784–789.

Massaro, D. W., Cohen, M. M., Campbell, C. S., and Rodriguez, T. (**2001**). "Bayes factor of model selection validates FLMP," Psychon. Bull. Rev. **8**, 1–17.

Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., and Tsuzaki, M. (**1993**). "Bimodal speech perception: An examination across languages," J. Phonetics **21**, 445–478.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.

Myung, I. J., and Pitt, M. A. (**1997**). "Applying Occam's razor in modeling cognition: A Bayesian approach," Psychon. Bull. Rev. **4**, 79–95.

Pitt, M. A., Kim, W., and Myung, I. J. (**2003**). "Flexibility versus generalizablity in model selection," Psychon. Bull. Rev. **10**, 29–44.

Pitt, M. A., and Myung, I. J. (**2002**). "When a good fit can be bad," Trends Cogn. Sci. **6**, 421–425.

Pugh, K. R., Shaywitz, B. A., Shaiwitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, R. T., Fletcher, J., Lacadie, C., Marchione, K., and Gore, J. C. (**1996**). "Auditory selective attention: An fMRI investigation," Neuroimage **4**, 159–173.

Robert-Ribes, J., Schwartz, J. L., and Escudier, P. (**1995**). "A comparison of models for fusion of the auditory and visual sensors in speech perception," Artif. Intell. Rev. **9**, 323–346.

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual and audio-visual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.

Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (**1997**). "The McGurk effect in infants," Percept. Psychophys. **59**, 347–357.

Rouger, J., (**2007**). "Perception audiovisuelle de la parole chez le sourd postlingual implanté cochléaire et le sujet normo-entendant: étude longitudinale psychophysique et neurofonctionnelle (Audiovisual speech perception in postlingually deaf cochlear-implant patients and normally-hearing subjects: Longitudinal psychophysical and neurofunctional study)," Ph.D. thesis, Université Paul Sabatier, Toulouse, France.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (**2007**). "Evidence that cochlear-implanted deaf patients are better multisensory integrators," Proc. Natl. Acad. Sci. U.S.A. **104**, 7295–7300.

Schorr, E. A., Fox, N. A., van Wassenhove, V., and Knudsen, E. I. (**2005**). "Auditory-visual fusion in speech perception in children with cochlear implants," Proc. Natl. Acad. Sci. U.S.A. **102**, 18748–18750.

Schwartz, J. L. (**2006**). "Bayesian model selection: The 0/0 problem in the fuzzy-logical model of perception," J. Acoust. Soc. Am. **120**, 1795–1798.

Schwartz, J. L., and Cathiard, M. A. (**2004**). "Modeling audio-visual speech perception. Back on fusion architectures and fusion control," in Proceedings of the ICSLP'2004, Jeju, Korea, pp. 2017–2020.

Schwarzer, G., and Massaro, D. W. (**2001**). "Modeling face identification processing in children and adults," J. Exp. Child Psychol. **79**, 139–161.

Seewald, R., Ross, M., Giolas, T., and Yonovitz, A. (**1985**). "Primary modality for speech perception in children with normal and impaired hearing," J. Speech Hear. Res. **28**, 36–46.

Sekiyama, K., and Burnham, D. (**2004**). "Issues in the development of auditory-visual speech perception: Adults, infants and children," in *Proceedings of the Eighth International Conference on Spoken Language Processing*, edited by S. H. Kim and D. H. Yuon (Sunjin Printing, Seoul, Korea), pp. 1137–1140.

Sekiyama, K., Burnham, D., Tam, H., and Erdener, D. (**2003**). "Auditory-Visual Speech Perception Development in Japanese and English Speak-

ers," in Proceedings of the International Conference Audio-Visual Speech Processing 2003, edited by J.-L. Schwartz, F. Berthommier, M.-A. Cathiard, and D. Sodoyer, St. Jorioz, France, pp. 43–47.
Sekiyama, K., and Tokhura, Y. (**1991**). "McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility," J. Acoust. Soc. Am. **90**, 1797–1825.
Sekiyama, K., and Tokhura, Y. (**1993**). "Inter-language differences in the influence of visual cues in speech perception," J. Phonetics **21**, 427–444.
Slate, E. H. (**1999**). "Assessing multivariate nonnormality using univariate distributions," Biometrika **86**, 191–202.
Strelnikov, K., Rouger. J., Lagleyre, S., Fraysse, B., Deguine, O., and Barone, P. (**2009**). "Improvement in speech-reading ability by auditory training: Evidence from gender differences in normally-hearing, deaf, cochlear implanted subjects," Neuropsychologia **47**, 972–979.
Teissier, P., Robert-Ribes, J., Schwartz, J. L., and Guérin-Dugué, A. (**1999**). "Comparing models for audiovisual fusion in a noisy-vowel recognition task," IEEE Trans. Speech Audio Process. **7**, 629–642.
Tiippana, K., Andersen, T. S., and Sams, M. (**2004**). "Visual attention modulates audiovisual speech perception," Eur. J. Cogn. Psychol. **16**, 457–472.
Walker, A. M. (**1967**). "On the asymptotic behaviour of posterior distributions," J. R. Stat. Soc. Ser. B (Methodol.) **31**, 80–88.
Welch, R. B., and Warren, D. H. (**1986**). "Intersensory interactions," in *Handbook of Perception and Human Performance, Volume I: Sensory Processes and Perception*, edited by K. R. Boff, L. Kaufman, and J. P. Thomas (Wiley, New York), pp. 25-1–25-36.

The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects

Kathryn Hopkins[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 13 February 2009; revised 4 December 2009; accepted 22 December 2009)

Speech reception thresholds (SRTs) were measured for target and competing-speech signals, processed to contain variable amounts of temporal fine structure (TFS) information. Signals were filtered into 30 1-ERB_N wide channels (where ERB_N refers to the bandwidth of normal auditory filters), which were either tone vocoded, preserving temporal envelope information (extracted using the Hilbert transform), or left unprocessed, containing both TFS and envelope information. Improvements in SRT were compared when TFS was progressively introduced, starting with the high- or low-frequency channels. The results suggest some redundancy in the TFS information across frequency regions. In a second experiment, the signal was divided into five, 6-ERB_N-wide spectral regions, four of which were tone vocoded. The remaining region was either absent (creating a spectral notch) or was present and unprocessed. SRTs were measured for normal-hearing and hearing-impaired subjects. Conditions where all channels were vocoded or unprocessed were also included. Normal-hearing subjects benefited similarly when TFS information was added to each region, suggesting that TFS information is important over a wide frequency range. Hearing-impaired subjects benefited less, although the benefit varied across subjects. Benefit from TFS information in speech was correlated with a psychophysical measure of TFS sensitivity obtained at two center frequencies. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3293003]

PACS number(s): 43.71.Es, 43.71.Ky, 43.66.Sr [RLF] Pages: 1595–1608

I. INTRODUCTION

Processing using a noise or tone vocoder can be used to remove temporal fine structure (TFS) information (the rapid fluctuations in amplitude close to the center frequency of a narrow-band signal) and to preserve, as far as possible, temporal envelope information (slower fluctuations in amplitude that are superimposed on this TFS). Speech reception thresholds (SRTs) for a target talker with a competing-talker or noise background are higher when the combined signal is noise or tone vocoded than when it is unprocessed. This is true even when such processing uses similar analysis bandwidths to those of the peripheral auditory filters, so that the spectral information available to the central auditory system is similar to that for unprocessed speech (Gnansia *et al.*, 2008; Hopkins *et al.*, 2008). This suggests that TFS information improves speech intelligibility for normal-hearing subjects when listening in background noise. Nonsense syllables processed to preserve TFS information and to remove temporal envelope information as far as possible are highly intelligible to normal-hearing subjects, following some training, suggesting that TFS may also play a role in speech intelligibility in quiet (Lorenzi *et al.*, 2006).

Recent studies suggest that subjects with moderate cochlear hearing loss have a reduced ability to use TFS information in psychophysical tasks (Buss *et al.*, 2004; Lacher-Fougère and Demany, 2005; Moore *et al.*, 2006a; Hopkins and Moore, 2007) and for the perception of speech (Buss *et al.*, 2004; Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). Hopkins and Moore (2007) measured the ability of normal-hearing and hearing-impaired subjects to discriminate harmonic and frequency-shifted tones, where each component was shifted upwards by the same amount in Hz. Such stimuli have the same envelope repetition rate, but different TFS. The harmonic and frequency-shifted tones were filtered into the same spectral region to prevent discrimination based on excitation-pattern cues, provided that components falling within the filter passband were unresolved. Normal-hearing subjects could discriminate such stimuli well when the filter was centered on the 11th component in the complex (and so all components within the passband were unresolved). For the same condition, hearing-impaired subjects performed very poorly, most scoring no better than chance. The authors interpreted this result as providing evidence that subjects with moderate cochlear hearing loss have a reduced ability to use TFS information.

Hopkins *et al.* (2008) measured the SRT as a function of the frequency range over which TFS information was available for a combined target and competing-talker signal. The signal was split into 32 channels, each of which was 1-ERB_N wide, where ERB_N refers to the equivalent rectangular bandwidth of the auditory filter for young, normal-hearing subjects, when measured at moderate sound levels (Glasberg and Moore, 1990). Above a cut-off channel, CO, channels were noise vocoded, and so contained only temporal enve-

[a]Author to whom correspondence should be addressed. Electronic mail: kh311@cam.ac.uk

lope information. For channels up to and including CO, channels were not processed, and so contained both TFS and envelope information. SRTs for normal-hearing subjects decreased as CO was increased, and more TFS information was included in the signal. As a group, the hearing-impaired subjects improved much less than the normal-hearing subjects as CO was increased, suggesting that they were less able to use TFS information in speech when listening in a competing-talker background. However, the individual results for the hearing-impaired subjects were rather variable, with some subjects showing no benefit at all from the added TFS information, and one subject benefiting almost as much as the normal-hearing subjects. The authors suggested that the reduced ability of hearing-impaired subjects to use TFS information could contribute to the speech perception problems reported by hearing-impaired subjects, and that the variability in the ability to benefit from TFS information could account for some of the variability in speech perception reported for subjects with similar audiometric configurations. However, the exact way in which TFS information contributes to speech intelligibility for normal-hearing subjects remains uncertain. A number of possibilities have been suggested, and these are described in Secs. I A–I C.

A. Fundamental frequency (F0) coding

Considerable research has addressed the importance of TFS information in pitch perception, particularly for coding the frequencies of pure tones and resolved harmonics in complex tones. Psychophysical data support the hypothesis that TFS information is important for the perception of pure tones for frequencies up to at least 5000 Hz. For example, Moore (1973) measured frequency difference limens (DLFs) for short-duration pure tones, which had a broader frequency spectrum than longer tones of the same frequency. DLFs were better than predicted by a model that assumed that discrimination of pure tones was based on detection of a difference in the excitation pattern (Zwicker, 1970), although others found smaller discrepancies between the predictions of Zwicker's model and DLFs as a function of duration (Freyman and Nelson, 1986). TFS information has also been suggested as a possible cue for coding F0 information for complex tones with components that are unresolved but intermediate in harmonic number (harmonic ranks 8–14) (Moore and Moore, 2003; Moore *et al.*, 2006b). However, this role remains uncertain, with other researchers suggesting that F0 coding for such stimuli can be accounted for by coding the repetition rate of the waveform using envelope information rather than TFS, or on the basis of partially resolved harmonics (Bernstein and Oxenham, 2006; Oxenham *et al.*, 2009). In the presence of a background sound, F0 information may be important for segregating sound sources into separate auditory streams, particularly when the background is a competing talker (Brokx and Nooteboom, 1982; Scheffers, 1983; Bregman *et al.*, 1990).

B. Coding the frequencies of peaks in the short-term speech spectrum

The coding of formant frequencies is important for vowel and consonant identification. Some animal studies have shown that a rate-place code may not carry sufficient information to account for psychophysical measures of vowel discrimination at all sound levels. At high levels, auditory filters broaden and the responses of many neurons saturate, so that peaks in the rate-place profile at the vowel formant frequencies become indistinct (Sachs and Young, 1979). However, Young and Sachs (1979) described representations of vowel formants using rate, place, and temporal information that were stable for levels up to at least 80 dB sound pressure level (SPL). These results suggest that temporal information may play some role in coding the frequencies of peaks in the spectra of speech signals.

C. Listening in the dips

It has been proposed that TFS information is important when listening to target speech in a background that is modulated in amplitude. In such a background, subjects can take advantage of a short-term improvement in signal-to-background ratio (SBR) to "glimpse" information about the target. Normal-hearing subjects have lower SRTs in modulated noise than steady noise (Duquesnoy, 1983; Peters *et al.*, 1998), and for a fixed overall SBR, they show better word and syllable identification for signals in modulated noise than in steady noise (Festen and Plomp, 1990). This improvement in performance has been termed "masking release." Hopkins and Moore (2009) measured SRTs for a target speaker in amplitude-modulated and steady noise. The target and background signals were mixed and then filtered into 32 1-ERB_N-wide channels. Channels up to and including a cut-off (CO) channel were not processed further, and so contained both envelope and TFS information, while channels above CO were tone vocoded to preserve only envelope information, which was extracted using the Hilbert transform. They found that normal-hearing subjects improved more when CO was increased (and so more TFS information was included) when listening in a modulated-noise background than in a steady-noise background. They suggested that subjects could use a change in TFS information in the dips of a modulated background to identify the presence of a signal and to code its properties. The idea that TFS information is important for masking release is supported by the positive correlation found by Lorenzi *et al.* (2006) between the ability of young hearing-impaired subjects to understand speech processed to contain primarily TFS information and masking release for intact speech in noise. However, it is known that, for stimuli processed to remove envelope cues, envelope information is partly reconstructed in the auditory system (Ghitza, 2001; Zeng *et al.*, 2004), and such reconstruction is greater when the filters in the peripheral auditory system are sharper. Hence, the correlation found by Lorenzi *et al.* (2006) might have occurred because frequency selectivity varied across subjects, and because frequency selectiv-

ity influences both masking release (Gnansia *et al.*, 2008) and the recovery of envelope cues from speech processed to contain primarily TFS information.

The interpretation of experiments on masking release has recently been addressed by Bernstein and Grant (2009). They showed that the amount of masking release depended on the SRT for steady noise; masking release was greater for conditions when the SRT for steady noise was more negative. This could partially explain the results of Lorenzi *et al.* (2006) and Hopkins and Moore (2009), as they measured masking release in terms of a reduction in SRT, and the SRT for steady noise was different across subjects (in the study of Lorenzi *et al.*) or conditions (in the study of Hopkins and Moore).

It has also been suggested that the coding of F0 information may be important for masking release (Summers and Leek, 1998; Stickney *et al.*, 2004). If this were the case, TFS information could be indirectly important for masking release because of its likely importance in coding F0 information (as discussed earlier). However, Oxenham and Simonson (2009) showed that masking release was similar for speech stimuli that were low-pass filtered at 1200 Hz or high-pass filtered at 1500 Hz. If accurate coding of F0 information was important for masking release, more masking release would be expected for the low-pass filtered condition, where resolved harmonics were present, allowing more accurate coding of the F0. Thus, these results suggest that F0 information may not be as important for masking release, as was previously thought.

D. Rationale for the present study

Measuring the importance of TFS in different spectral regions may provide information about the possible role of TFS information in speech perception, and the problems that may be encountered by people who are relatively insensitive to TFS, such as hearing-impaired people and cochlear-implant users. If TFS information is mainly important for coding the frequencies of resolved harmonics (which dominate the perception of F0 information), then a benefit from TFS information would only be expected for frequencies below about 1000 Hz. If, however, TFS information is also important for coding peaks in the short-term spectrum of speech, or for listening in the dips of a fluctuating background, then benefit from TFS information should occur for higher frequencies, provided that the listener has the ability to extract the TFS information.

Hopkins *et al.* (2008) found that, when TFS information was added starting from the low-frequency channels, the biggest improvement in performance occurred when CO was increased from 0 to 4, and the improvement as CO was raised further was gradual. No significant improvement in performance occurred when CO was increased beyond 24 (equivalent to a frequency of 4102 Hz). It may initially appear that this supports the hypothesis that TFS information is mainly important at low frequencies, and that TFS information is not at all important for frequencies above about 4000 Hz. However, this result could be influenced by redundancy in the TFS information that was added to the signal as CO was increased. If there was redundancy in the TFS information in different spectral regions, adding more TFS information would not necessarily result in an improvement in performance.

This hypothesis was tested in experiment 1. A similar method to that of Hopkins *et al.* (2008) was used, but in addition to the TFS information being added starting with the low-frequency channels (TFS-low conditions), TFS information was added starting with the high-frequency channels (TFS-high conditions). If the benefit from adding TFS information in a particular spectral region is reduced by the presence of TFS information in other spectral regions (because of redundancy in the TFS information), then the changes in performance between the same values of CO for the TFS-low and TFS-high conditions may be different. Only normal-hearing subjects were tested.

In experiment 2, the importance of TFS information in different spectral regions was assessed in more detail by measuring the benefit of adding TFS information to single isolated spectral regions to reduce the effects of redundancy in the TFS. Normal-hearing and hearing-impaired subjects were tested.

In experiment 3, a psychophysical measure of TFS sensitivity developed by Moore and Sek (2009a) was used for the same hearing-impaired subjects who took part in experiment 2 and for a different group of normal-hearing subjects. Subjects were required to discriminate harmonic complex tones from frequency-shifted tones, in which all of the components were shifted up by the same amount in Hz. The stimuli were centered on either 1419 or 2904 Hz, which corresponded to the third and fourth frequency regions in experiment 2. It was assessed whether the results of this measure of sensitivity to TFS were correlated with the benefit obtained from TFS information in speech. If this were the case, it would lend weight to the test as a measure of sensitivity to TFS, and also to the idea that the ability to use TFS is an important predictor of the ability of hearing-impaired subjects to understand speech in complex backgrounds.

II. METHOD

A. Subjects and materials

Seven normal-hearing subjects were tested. All had thresholds of 15 dB hearing level (HL) or better in their test ear at the standard audiometric frequencies from 250 to 8000 Hz, and were aged between 19 and 24 years. All subjects were native British English speakers.

Sentences from the coordinate response measure (CRM) corpus were used for testing (Moore, 1981; Kitterick and Summerfield, 2007). CRM sentences all have the same standard format: "Ready *callsign*, go to *color*, *number* now!" Subjects were required to identify the callsign, color, and number that they heard from a list of four (in the case of color) or eight (in the case of callsign and number) alternatives. Sentences were presented in a competing-talker background. A competing-talker background was used because, in a previous study, this background gave a large reduction in SRT when TFS information was added (Hopkins *et al.*, 2008). The background was randomly selected from a pas-

sage of continuous prose, spoken by a different talker. Silent intervals between sentences and pauses for breath were removed from the passage by hand editing. Natural-sounding pauses between words of less than 200 ms were left intact. Both target and competing talkers were male speakers of British English. The F0 of the target talker's voice was estimated to range from 130 to 200 Hz, while the competing talker had a larger F0 range of about 130 to 280 Hz.

B. Processing and equipment

Target and competing-speech signals were added together and subsequent processing was carried out on the combined signal. The signal was filtered into 30 channels, spanning a frequency range of 100 to 8000 Hz with an array of linear-phase, finite-impulse-response (FIR) filters. The filters had a variable order, chosen such that the transition bands of each filter had similar slopes when plotted on a logarithmic frequency scale. Each filter was designed to have a response of −6 dB (relative to the peak response) at the frequencies at which its response intersected with the responses of the two adjacent filters. Each channel was 1-ERB_N wide. Hence, the processing was not expected to have a large effect on the spectral information available to the central auditory system. Channels were divided into two groups by a cut-off channel, CO. For TFS-low conditions, channels up to and including CO were not processed further, and channels above CO were tone vocoded. For TFS-high conditions, channels above CO were not processed further, and channels up to and including CO were tone vocoded. For channels that were vocoded, the Hilbert transform was used to find the "analytic signal" for each channel (Hilbert, 1912; Bracewell, 1986), and the envelope was extracted by finding the absolute value of this analytic signal. Each channel envelope was used to modulate a sine wave with a frequency equal to the center frequency of the channel. Each modulated sine wave was then filtered using the original channel analysis filters to remove side bands that fell outside of the channel bandwidth. Finally, signals from all channels were aligned in time and combined.

All signals were generated with a high-quality 16-bit PC soundcard (Lynx One) at a sampling rate of 22 050 Hz, passed through a Mackie 1202-VLZ mixing desk, and presented to the subject monaurally via Sennheiser HD580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

C. Procedure

Six values of CO were used (0, 6, 12, 18, 24, and 30) for both the TFS-low conditions and the TFS-high conditions. For the TFS-low conditions, unprocessed information was included in channels up to and including CO (as in Hopkins *et al.*, 2008), and for the TFS-high conditions, unprocessed information was included in channels above CO. Note that the TFS-low condition with CO=0 was equivalent to the TFS-high condition with CO=30, and the TFS-high condition with CO=0 was equivalent to the TFS-low condition with CO=30. These were the conditions for which the whole signal was vocoded (the former) or unprocessed (the latter). All other values of CO were tested for both TFS-low and TFS-high conditions, giving ten conditions in total.

SRTs corresponding to the 70.7% correct point on the psychometric function were measured for each subject and condition. At the start of each run, the SBR was set to a value for which the subject was expected to identify all of the words in the target sentence correctly. A sentence was considered correct if the callsign, color, and number were correctly identified, and incorrect if any mistakes were made. A two-up, one-down adaptive procedure was used to set the SBR for each trial. If two sentences were correctly identified in succession, the SBR for the next sentence was set k dB lower; if a sentence was incorrectly identified, the next sentence was presented at a SBR that was k dB higher. For the first two turnpoints, k was equal to 4 dB, and for the remaining six turnpoints, k was equal to 2 dB. The SRT was taken to be the mean of the SBR values at the last six turnpoints. For SBRs of −16 dB or higher, the level of the target talker was fixed at 65 dB SPL, and the level of the competing talker was varied. For SBRs lower than −16 dB, the level of the competing talker was not increased further, to prevent the level of the combined signal from becoming uncomfortably loud; instead, the level of the competing talker was fixed, and the level of the target talker was reduced to give the desired SBR.

D. Analysis

A concern with the methods used here (and by Hopkins *et al.*, 2008) is the extent to which the tone-vocoder processing is successful in preserving the temporal envelope information available to the central auditory system while removing the TFS information. Here, the tone-vocoder processing was carried out using channels that were 1-ERB_N wide. This channel width was chosen to maximize the fidelity of envelope transmission to the central auditory system, while ensuring that TFS information was disrupted; the fidelity of envelope transmission is expected to increase as narrower channels are used, but this may also lead to a crude representation of the TFS in the tone-vocoded signal. To assess the effects of peripheral filtering on the information available to the central auditory system, both a tone-vocoded signal and the corresponding original signal were passed through an array of bandpass filters that were designed to have similar frequency selectivity to normal auditory filters (Glasberg and Moore, 1990). Fifty-nine overlapping fourth-order gammatone filters were used, spaced between 100 and 8000 Hz. Each had a bandwidth of 1-ERB_N. At the output of each filter, the envelope and TFS were decomposed using the Hilbert transform. The correlation, $r$env, between the envelope produced by the original signal and that produced by the tone-vocoded signal was found for each channel. Similar correlations were obtained for the TFS, giving rTFS. The original and tone-vocoded signals were time aligned, prior to measuring the correlations to compensate for the delays introduced by the tone-vocoder processing. Small misalignments in the original and processed signals could lead to a much-reduced value of rTFS, as the TFS within each channel fluctuates rapidly. To check that the time alignment had

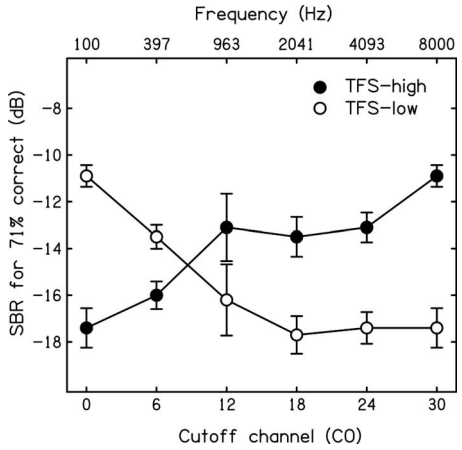


FIG. 1. SRTs plotted as a function of CO for TFS-high and TFS-low conditions. Error bars are plotted at ±1 standard error of the mean.

been successful, the correlations were measured with the processed signal advanced or delayed by time intervals up to 1 period of the channel center frequency. Correlations were measured for 20 time intervals in this range, and *r*TFS was taken to be the largest correlation measured. Averaged across the 59 filters, *r*env was equal to 0.936 and *r*TFS was equal to 0.014, indicating that the tone-vocoder processing was relatively successful at preserving the temporal envelope information and removing the TFS information from the original signal. There was a small difference between the value of *r*env for filters that had the same center frequency as the channel center frequencies used for the tone-vocoder processing, and the value of *r*env for the filters with center frequencies half way between those for the tone vocoder. The mean values of *r*env were 0.941 for the former and 0.932 for the latter. The same analysis was also carried out with an array of 59 filters with the same center frequencies, but with bandwidths of 2 ERB_N to simulate the loss of frequency selectivity that is typically found for subjects with moderate cochlear hearing loss (Glasberg and Moore, 1986). The mean value of *r*env at the output of these filters was 0.936, and the mean value of *r*TFS was 0.014. Thus, the vocoder processing was successful in preserving envelope cues and removing TFS cues when the simulated auditory filters were broader than normal.

III. RESULTS

SRTs as a function of CO are plotted in Fig. 1. For the TFS-low condition, where unprocessed information was present in channels up to and including CO, as in Hopkins *et al.* (2008), performance improved as CO increased, and more TFS information was presented, starting from the low-frequency channels. A one-way, repeated-measures, analysis of variance (ANOVA) showed that there was a significant effect of CO [$F(5,30)=12.3$; $p<0.001$]. A Fisher's least significant difference (LSD) procedure was used to test whether SRTs for different values of CO were different from each other. The results are shown in the top panel of Table I. SRTs improved for values of CO up to 12, but there was no significant improvement for higher values of CO.

TABLE I. Differences between mean SRTs measured with different pairs of values of CO for TFS-high and TFS-low conditions. Equivalent cut-off frequencies (Hz) are also shown. The LSDs calculated using Fisher's procedure were 2.3 (TFS-low condition) and 2.1 (TFS-high condition). Differences equal to or above this value are shown in bold.

		TFS-low					
CO		0	6	12	18	24	30
	Frequency	100	397	963	2041	4093	8000
0	100	0					
6	397	**2.7**	0				
12	963	**5.3**	**2.6**	0			
18	2041	**6.8**	**4.1**	1.5	0		
24	4093	**6.5**	**3.8**	1.2	0.3	0	
30	8000	**6.5**	**3.9**	1.2	0.3	0.0	0
		TFS-high					
CO		0	6	12	18	24	30
	Frequency	100	397	963	2041	4093	8000
0	100	0					
6	397	1.3	0				
12	963	**4.3**	**2.9**	0			
18	2041	**3.9**	**2.6**	0.4	0		
24	4093	**4.3**	**2.9**	0.0	0.4	0	
30	8000	**6.5**	**5.2**	**2.3**	**2.7**	**2.3**	0

For the TFS-high condition, performance improved as CO was reduced, and more TFS information was presented, starting from the high-frequency channels. The effect of CO was significant [$F(5,30)=10.03$; $p<0.001$]. LSD test results are shown in the bottom panel of Table I. The pattern of the results broadly mirrored that for the TFS-low condition, with most change in performance when TFS was added to the low-frequency channels. However, performance significantly improved even when TFS information was added to the highest-frequency channels, a result that was not found for the TFS-low condition or by Hopkins *et al.* (2008).

IV. DISCUSSION

For both the TFS-low and TFS-high conditions, performance improved as more TFS information was added to the signal. For the TFS-low condition, performance improved as CO was increased. This is broadly consistent with the results of Hopkins *et al.* (2008), although Hopkins *et al.* (2008) found significant improvements for higher values of CO (up to CO=24 rather than up to CO=12), and the improvement in SRT from adding TFS across the whole spectrum was much larger (15.5 dB compared with 6.5 dB). A number of factors could account for these differences. First, Hopkins *et al.* (2008) used a noise vocoder, whereas a tone vocoder was used here. Whitmal *et al.* (2007) compared the intelligibility of noise and tone-vocoded stimuli, and found that the noise-vocoded stimuli were less intelligible, particularly when very narrow bands were used, as was the case for the bands used here and by Hopkins *et al.* (2008). Similarly, Stone *et al.* (2008) found that intelligibility was significantly higher for a tone vocoder than a noise vocoder when listening to speech in a competing-talker background. Whitmal *et al.* (2007) suggested that this difference in performance arose because

the narrow-band noise carriers used for the noise vocoder introduced random modulations, which could be confused with the envelope modulations in the speech signal. A tone carrier does not introduce these spurious modulations in the envelope of the vocoded signal. Distortions introduced by the noise vocoder could partly account for the poorer performance when CO was low, and a large proportion of the signal was vocoded. Souza and Rosen (2009) also pointed out that a tone vocoder may better preserve high-rate periodicity information, due to the presence of evenly spaced components corresponding to the side bands of a tone carrier.

For both the TFS-low and TFS-high conditions, the greatest changes in SRT occurred when TFS information was added to channels with center frequencies below 1000 Hz, which is consistent with the idea that TFS information is important for coding F0 information. However, performance was significantly better when CO=24 than when CO=30 for the TFS-high condition. For the TFS-low condition, there was no significant difference in performance when CO=24 and CO=30. This suggests that there is some redundancy in TFS information across spectral regions; adding TFS information to the highest-frequency region only improved performance when the rest of the channels were vocoded (in the TFS-high condition). When the other channels were unprocessed (as in the TFS-low condition), no improvement occurred, perhaps because the TFS information in that region was providing information that was redundant with the TFS information in lower-frequency channels; as a result, TFS information in the high-frequency channels did not add any information to that already present in the low-frequency channels.

The redundancy in the TFS information across frequency might be explained by the use of TFS information for "listening in the dips." TFS information in a narrow spectral region could be used to identify signal portions in the dips of a fluctuating masker. Envelope information in speech is highly correlated across frequency regions (Steeneken and Houtgast, 1999), so this information could be used to identify speech portions with a favorable SBR across a broad spectral region. The redundancy in TFS information could also be explained by a general redundancy in speech information across spectral regions. Turner *et al.* (1998) filtered nonsense syllables into four frequency bands, and measured weightings for the importance of each band for intelligibility. They found that the weightings differed depending on the combination of bands that was presented, suggesting some redundancy across bands. If TFS information is important for carrying phonetic information, then a general redundancy in speech information across frequency could explain the redundancy in TFS information that we observed.

Some aspects of the data from experiment 1 could be accounted for if it were assumed that the SRT had a "floor" value of about −18 dB. However, it is not clear why this should be the case; there was no limit imposed on the minimum SBR that could be presented, and information in the dips of a highly modulated masker (such as the competing talker used here) would be available at very negative SBRs. Indeed, two of the individual normal-hearing subjects had SRTs of less than −20 dB for the condition in which all of the channels were unprocessed. This floor in performance of −18 dB coincides with the SBR at which the signal was attenuated rather than the masker made louder to give the required SBR that was dictated by the adaptive procedure (see Sec. II C for details). This was done to prevent the signal from becoming uncomfortably loud. One possibility is that the target became less audible below this SBR, and so it contributed to a floor effect. However, this explanation is unlikely, as the target would have still been well above threshold for normal-hearing subjects down to much more negative SBRs than were used here.

V. EXPERIMENT TWO

A. Rationale

The results of experiment 1 suggest that there is some redundancy in TFS information across spectral regions. This means that it is difficult to infer the relative importance of TFS information in different spectral regions from data where TFS information has been added to a speech signal cumulatively (as was the case in experiment one and in Hopkins *et al.*, 2008), as the improvement in performance as TFS is added depends not only on the spectral region to which the information is added, but also on the TFS information that is already present in the signal. In experiment 2, to partially overcome this problem, the benefit of adding TFS to single isolated spectral regions was assessed. Both normal-hearing and hearing-impaired subjects were tested. Hopkins and Moore (2007) found that subjects with moderate cochlear hearing loss were often completely unable to use TFS information in complex sounds to discriminate harmonic and frequency-shifted tones. However, only tones with components above about 1 kHz were tested, and so it is possible that hearing-impaired subjects are generally more sensitive to TFS information at lower frequencies. Hearing-impaired subjects were tested in experiment 2 to investigate whether their reduced benefit from TFS information in speech, as measured by Hopkins *et al.* (2008) when TFS information was added starting from the low frequencies, was due to a general limitation in processing TFS information or due to a limitation specifically in processing TFS information at higher frequencies.

Another concern is that the improvement in performance when TFS information is added to a particular spectral region could be affected by the distribution of speech information across the frequency spectrum. For example, if TFS was added to a region that contributed little to speech intelligibility, a smaller improvement in performance might be expected than if TFS information was added to a spectral region that contributed a lot to speech intelligibility, even if TFS information was equally available in the two cases. This is a particular concern when testing hearing-impaired subjects, as access to TFS information in a particular spectral region could be limited by the audibility of the signal in that region, even when frequency-dependent gains are applied to try to ensure audibility, as was the case here (see below). To assess the contribution of each spectral region to speech intelligibility, the importance of envelope information within each spectral region was assessed by comparing performance

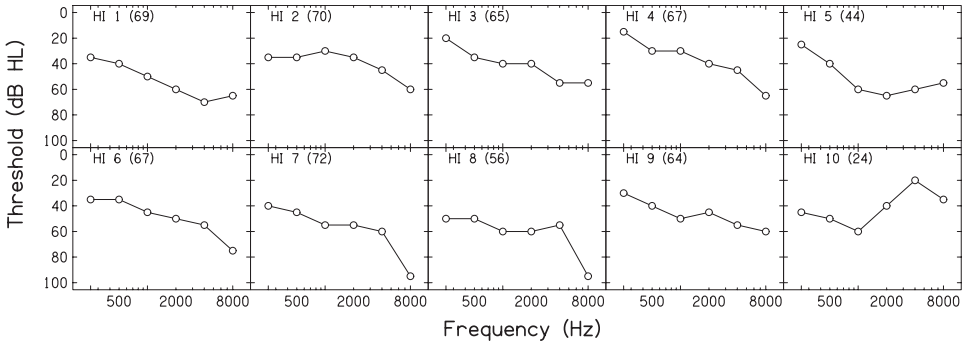


FIG. 2. Audiometric thresholds of the test ears of the hearing-impaired subjects. The age of each subject is shown in parentheses.

when a spectral notch was present in each frequency region (and all other regions were vocoded) to performance when all of the spectral regions were tone vocoded.

The frequencies at which TFS information is important would be expected to be different, depending on the type of information that is coded by TFS. If TFS information is primarily used for coding F0, TFS information should mainly be important for frequencies below about 1000 Hz, where resolved harmonics are present for the F0 values that typically occur in speech (although some information about the F0 would be present at higher frequencies, in the form of unresolved harmonics). If TFS information is used for coding information about the frequency spectrum of a sound, or for listening in the dips of a fluctuating masker, then TFS information could be important for frequencies above 1000 Hz, provided that the TFS information at these frequencies can be used by the central auditory system.

B. Methods

1. Subjects

The seven normal-hearing subjects who took part in experiment 1 also took part in experiment 2. Ten hearing-impaired subjects also took part. The audiometric thresholds of the test ears of the hearing-impaired subjects are shown in Fig. 2. The age of each subject is shown in parentheses. The better-hearing ear was used as the test ear for all subjects, apart from HI 10. The worse-hearing ear was used for HI 10 because this ear had previously been used for other psychoacoustic experiments in the laboratory. For the hearing-impaired subjects, the TEN(HL) test described by Moore *et al.* (2004) was used to assess whether cochlear dead regions were present. The results for the hearing-impaired subjects gave no evidence for cochlear dead regions between 500 and 4000 Hz in their test ears. None of the hearing-impaired subjects had air-bone gaps greater than 15 dB between 500 and 4000 Hz, indicating that their hearing losses were sensorineural in origin.

2. Stimuli, processing, and equipment

As for experiment 1, the target and background talker signals were summed, prior to further processing. The combined signal was filtered into 30 channels, as before, and the channels were divided into five spectral regions, each 6-ERB_N wide. The channel signals for four of the spectral regions were tone vocoded. For the remaining spectral region (1–5), two conditions were tested: one where the channel signals from that region were discarded (− conditions, 1−, 2−, 3−, 4−, and 5−), leading to a spectral notch in the processed signal, and one where there was unprocessed information in that spectral region (+ conditions, 1+, 2+, 3+, 4+, and 5+). For the hearing-impaired subjects, two comparison conditions were included: a condition where all channels were tone vocoded (allvoc), and a condition were all channels were unprocessed (allunproc). These conditions had already been included for the normal-hearing subjects, as part of experiment 1. For the normal-hearing subjects, as in experiment 1, the target speech was presented at a level of 65 dB SPL for SBRs of −16 dB or higher. The level of the competing talker was set to give the desired SBR. For SBRs lower than −16 dB, the level of the competing talker was not increased further, to prevent the combined signal from becoming uncomfortably loud. Instead, the level of the target talker was reduced to give the desired SBR. For the hearing-impaired subjects, gains were applied to the combined signal, according to the "Cambridge formula" hearing aid fitting prescription, individually according to the audiogram for each hearing-impaired subject (Moore and Glasberg, 1998). Gains were specified at audiometric frequencies between 250 and 6000 Hz. The Cambridge formula gains are designed to ensure speech audibility between these frequencies. The sentence material and test equipment were the same as for experiment 1.

3. Procedure

The procedure was the same as for experiment 1, except that the hearing-impaired subjects completed five runs per condition rather than three, so that the results for the different conditions could be compared more accurately for each subject, since previous studies have shown large individual differences among hearing-impaired subjects in their ability to use TFS (Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). For the normal-hearing subjects, experiments 1 and 2 were run simultaneously. Conditions from experiments 1 and 2 were interleaved during testing, and in cases where conditions for the two experiments were the same (for example the TFS-low CO=6 condition in experiment 1 and the 1+ condition in experiment 2), testing was not duplicated.

C. Results

Table II shows the mean scores [with corresponding standard deviations (SDs)] for the two comparison conditions and for each of the + and − conditions. The benefit

TABLE II. Mean and SDs of the SRTs for the normal-hearing and hearing-impaired subjects for experiment 2.

	Normal-hearing		Hearing-impaired	
	Mean	SD	Mean	SD
allunproc	−17.4	2.0	−5.8	6.3
allvoc	−10.9	1.1	−3.3	5.4
1+	−13.5	1.2	−3.8	6.2
2+	−14.0	1.6	−4.3	6.1
3+	−13.6	2.6	−3.7	5.2
4+	−12.7	1.5	−3.6	4.9
5+	−13.1	1.5	−3.2	4.9
1−	−8.6	2.2	0.9	4.4
2−	−7.5	2.8	1.3	5.0
3−	−6.4	2.5	−0.1	4.1
4−	−10.2	2.3	−1.9	4.8
5−	−10.0	3.5	−2.3	3.9

(improvement in SRT) from adding envelope information (open circles in Figs. 3 and 4) was calculated as the difference between the SRT for the allvoc condition and the SRT for the − condition for each spectral region. Similarly, the benefit from adding envelope and TFS information together (filled circles in Figs. 3 and 4) was calculated as the difference between SRTs for the − and the + conditions for each spectral region. The mean benefit scores for the normal-hearing subjects are shown in Fig. 3. A two-way within-subjects ANOVA on these scores showed significant effects of both spectral region (1–5) and information type (envelope and envelope+TFS) [$F(4,24)=4.65$; $p=0.006$ and $F(1,6)=6.46$; $p=0.04$, respectively], but no significant interaction between the factors [$F(4,24)=0.66$, $p=0.63$], suggesting that the relative benefit of TFS information was similar for each spectral region.

Mean benefit scores for the hearing-impaired subjects are shown in Fig. 4. A two-way within-subjects ANOVA on the mean benefit scores for the hearing-impaired subjects showed that there was a significant effect of spectral region [$F(4,36)=6.41$; $p<0.001$], but no significant effect of information type [$F(1,9)=0.03$; $p=0.87$], suggesting that, as a group, there was no significant benefit of TFS information over that provided by envelope information. There was no significant interaction between spectral region and the type of information that was added [$F(4.36)=1.11$; $p=0.37$]. However, there was a large amount of variability in the pattern of results across hearing-impaired subjects.

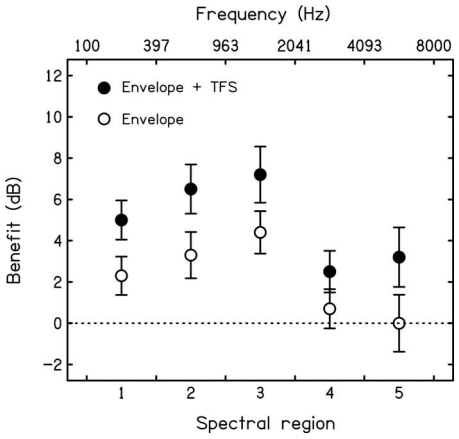


FIG. 3. Mean results of experiment 2 for the normal-hearing subjects. The benefit of adding envelope information (open circles) was calculated as the difference between the SRT for the allvoc condition and the SRT for the − condition for each spectral region. The benefit from adding envelope and TFS information together (filled circles) was calculated as the difference between SRTs for the corresponding − and + conditions. Error bars are plotted at ±1 standard error of the mean.

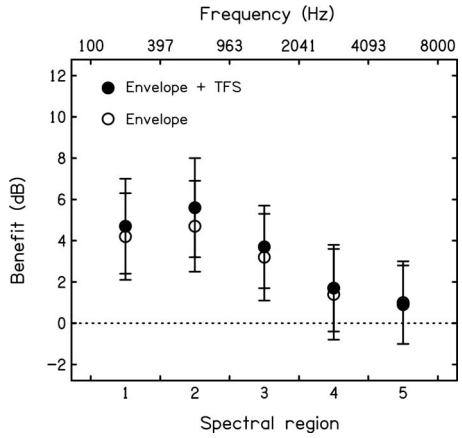


FIG. 4. As Fig. 3, but for the hearing-impaired subjects.

The individual results are shown in Fig. 5. Variability within individual subjects was also high, despite five SRT measurements per condition. Mean standard errors for the SRT measurements for individual subjects ranged from 0.6 to 1.5 dB. Some subjects showed a similar pattern of results to the normal-hearing subjects (e.g., HI 6 and HI 3, except for region 4), whereas other subjects showed no benefit from addition of TFS in most spectral regions (e.g., HI 1 and HI 4). A separate two-way ANOVA was performed on the benefit scores for each hearing-impaired subject with factors spectral region and information type. The benefit score for the addition of TFS information to a given spectral region was obtained as the difference between scores for the + and − conditions for that region. Even though there were five measures for each condition, the repetitions were not paired. Therefore, the difference between the mean scores for the five estimates in the + and − conditions was used in the ANOVA. This meant that there were no repeated measures. Hence, the interaction term was used to estimate the residual variance in the ANOVA. The effect of spectral region was significant ($p<0.05$) for subjects HI 1, HI 2, HI 4, HI 5, HI 6, HI 7, and HI 10, but the effect of information type was significant only for HI 6.

The mean benefit of adding TFS to the whole frequency spectrum (performance for the allunproc condition relative to performance for the allvoc condition) was 6.5 dB for the normal-hearing subjects, and 2.5 dB for the hearing-impaired subjects. A two-way, mixed design ANOVA with a between-subjects factor of subject type (normal-hearing and hearing impaired) and a within-subjects factor of condition (allvoc and allunproc) was performed on the SRTs for all subjects for the two conditions. The effects of subject type and condition were both significant [$F(1,15)=16.7$; $p<0.001$ and

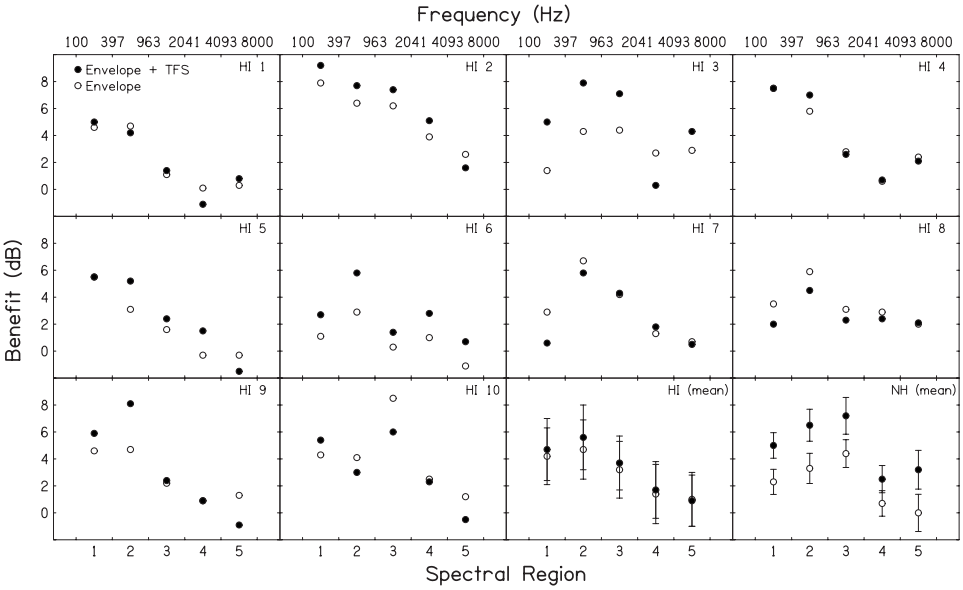


FIG. 5. Individual results of experiment two for the hearing-impaired subjects. Benefit scores were calculated in the same way as for Fig. 3. The bottom-right panels show the mean results for the hearing-impaired and normal-hearing subjects, for comparison. Error bars are plotted at ±1 standard error of the mean.

$F(1,15)=43.8$; $p<0.001$, respectively], and there was a significant interaction between the two factors [$F(1,15)=10.0$; $p=0.006$]. The individual scores for the hearing-impaired subjects are presented, together with the results of experiment 3, in Sec. VI B.

D. Discussion

For the normal-hearing subjects, the similar benefit of adding TFS information to the already-present envelope information for each of the spectral regions tested suggests that the importance of TFS information is not restricted to the coding of F0. If TFS information was only important for coding the F0s of the target and competing talkers, the most benefit would be expected for spectral regions 1 and 2, where resolved harmonics would have been present. The results are consistent with the idea that TFS information is important across the frequency spectrum. TFS could be important for coding the frequencies at which there are spectral peaks. For example, if at a particular time, there is a formant centered at frequency *fx* (and hence, one or more relatively intense stimulus components near *fx*), then, provided that these stimulus components are unresolved, channels centered close to *fx* will show TFS synchronized to *fx*, and this will be reflected in the patterns of phase-locking in those channels. Physiological evidence suggests that there may be insufficient information in the rate-place code to represent vowel formants at high levels (Young and Sachs, 1979). If information from the timing of spike discharges is included, then the representation of vowel formants is stable over a wide range of sound levels (Young and Sachs, 1979).

TFS information may also be important for conveying information about the quality of sounds. For example, a fricative such as *s* or *sh* has energy at high frequencies that is noisy. Replacing this noisy TFS information with a tone, as occurs with tone-vocoder processing, may impair the perception of fricatives, if information about the quality of the sound (its noisiness in this case) is carried by TFS information.

Another possible role for TFS information is related to listening in the dips. Hopkins and Moore (2009) showed that normal-hearing subjects benefited more from TFS information when listening in modulated noise than in steady noise. In the present study, TFS information could have been used for dip listening at any frequency region, provided the TFS information could be used by the central auditory system.

The benefit from TFS information at high frequencies (above 4000 Hz) is somewhat surprising, as it has traditionally been assumed that phase-locking information above 4000–5000 Hz is unusable. This assumption comes from physiological data in mammalian species other than humans, showing that phase locking to pure tones above this frequency is much degraded. Additionally, human psychophysical measurements that are assumed to reflect temporal mechanisms at low frequencies often worsen above 4000–5000 Hz (for example, pure-tone frequency discrimination), suggesting that temporal information is badly preserved at such frequencies. However, the limit of phase locking in humans is unknown. Heinz *et al.* (2001) used an auditory nerve model to investigate whether human pure-tone frequency and level discrimination could be accounted for by rate-place information alone at high frequencies. They used filters that had the same properties as human auditory filters (measured psychophysically) and phase-locking characteristics that were the same as those measured in the cat. They found that psychophysical performance was best predicted when both rate-place and temporal information were included in the model, for frequencies up to at least 10 000 Hz. This result suggests that, even if phase-locking characteristics in humans are similar to those in other mammalian species, some TFS information could be useful even at very high frequencies.

This hypothesis is supported by the data of Moore and Sek (2009b), who showed that normal-hearing subjects could discriminate complex tones on the basis of their TFS for frequencies up to at least 8000 Hz. They measured the discrimination of harmonic and "frequency-shifted" stimuli (where each component was shifted upwards by the same

amount in Hertz) that were filtered into the same frequency region to reduce excitation-pattern cues. They used the same method, as described in Sec. VI A (see later), but with high F0s (800 and 1000 Hz) and with the bandpass filter centered at 14F0 so that all of the components in the complexes fell above 8000 Hz. Moore and Sek (2009b) suggested that subjects may be able to use TFS derived from unresolved harmonics of complex tones at frequencies for which coding of pure tones using TFS is much reduced. They argued that, at high frequencies, the decoding of phase locking may be limited by refractoriness in the responses of auditory nerve fibers, because the refractory period of nerve fibers would be long relative to the time between peaks in the stimulus waveform. This means that the responses of many auditory nerve fibers would need to be compared to determine the frequency of a pure tone, because a single nerve fiber would be in a refractory state for many periods of the waveform. The situation is different for a periodic complex tone with the same center frequency that contains only harmonics that are unresolved. In this case, nerve fibers probably only fire in response to a few TFS peaks that are close to maxima in the envelope of the waveform. It is assumed that performance is based on time intervals between peaks in the TFS close to adjacent envelope maxima. These time intervals are much longer than the intervals between adjacent peaks in the TFS. Coding of these longer time intervals would be much less affected by refractoriness in auditory nerve fibers.

The reason that TFS information resulted in a reduction in SRT when added to the highest-frequency region alone could be related to the role of TFS information in conveying the noisy quality of the high-frequency energy in fricative sounds. The same result might not have been found if a noise vocoder had been used, rather than a tone vocoder, as in this case, the TFS information in the high-frequency channels would have retained a noisy quality when vocoded.

A possible concern with the methods used here is that the tone-vocoder processing introduces distortion in the envelope that could account for the poorer performance when the signals were entirely vocoded than when some unprocessed information was present. However, it seems unlikely that this could account for all of the differences in SRT that were observed. As described earlier, the correlations between the envelopes at the outputs of simulated auditory filters, in response to unprocessed and tone-vocoded stimuli, were high (over 0.93), and were similar when the simulated filters were 1-ERB_N or 2-ERB_N wide (simulating normal and impaired frequency selectivity, respectively). Given this, it is not clear why normal-hearing listeners would be affected differently than hearing-impaired listeners by the envelope distortions, and so it seems unlikely that this could account for the different pattern of results between the two groups. Also, for the highest spectral region (region 5), normal-hearing subjects showed almost no benefit from envelope information alone, but a large benefit when TFS and envelope information was added. It seems unlikely that this relatively large improvement could be accounted for by a slight improvement in envelope fidelity, given that envelope information alone had almost no effect on the SRT. However, disruption of envelope cues produced by the tone vocoder could account for some of the benefits that we have ascribed to the use of TFS information. Specifically, when TFS information was added in a particular spectral region (so that the signal was intact in that region), the envelope would have been coded somewhat more faithfully.

The results for the hearing-impaired subjects are consistent with those of Hopkins *et al.* (2008); the hearing-impaired subjects benefited less than the normal-hearing subjects from TFS information in speech, both when it was added to the whole spectrum, and when it was added to a narrow spectral region. The benefit to the hearing-impaired subjects from the addition of envelope information alone was similar to or larger than that found for normal-hearing subjects, which suggests that the gains that were applied to the signal, according to the audiogram of each subject, were successful in restoring audibility. The variability in the results across hearing-impaired subjects is also consistent with the findings of Hopkins *et al.* (2008), although the results for individual subjects were also rather variable, and so it was difficult to draw conclusions from the patterns of results for individual subjects. The lack of a significant benefit from the addition of TFS information for most of the hearing-impaired subjects could be due to the large standard errors associated with each benefit score. The differences across individual hearing-impaired subjects are discussed further in Sec. VI C.

VI. EXPERIMENT THREE: MEASURING SENSITIVITY TO TFS

The results of experiment 2 showed large individual variability in the ability of hearing-impaired subjects to benefit from TFS information in speech, when TFS information was added to the whole frequency spectrum. In experiment 3, sensitivity to TFS was measured for the same hearing-impaired subjects, using a psychophysical procedure. The results of the two experiments were compared to investigate the relationship between sensitivity to TFS in complex tones and the ability to use TFS information in speech in a competing-talker background.

A. Methods

The "TFS1" test described by Moore and Sek (2009a) was used, which is based on the method described by Hopkins and Moore (2007). The ten hearing-impaired subjects who took part in experiment 2 also took part in experiment 3. Four normal-hearing subjects (who did not take part in experiments 1 or 2) were also tested. The normal-hearing subjects were aged between 21 and 24 yrs and none were musically trained. The TFS1 test was designed as a quick method for measuring TFS sensitivity. The procedure took approximately 1 h, including task familiarization and practice runs. The same equipment was used as for experiments 1 and 2.

1. Stimuli

Subjects were required to discriminate harmonic tones (H) from frequency-shifted tones (S), in which each component was shifted upwards by the same amount in Hz(δ). Such complexes have the same envelope repetition rate, but differ-

ent TFS. Components were filtered using a fixed passband to minimize differences in the excitation patterns evoked by the H and S tones. The passband of the filter was 5F0 wide, and the skirts of the filter had a slope of 30 dB/octave. Two F0s were used (129 Hz and 264 Hz), and the bandpass filter was centered on the 11th harmonic in each case (corresponding to center frequencies of 1419 and 2904 Hz, respectively). These center frequencies were chosen to be the same as the center frequencies of spectral regions 3 and 4 in experiment 2. Center frequencies corresponding to spectral regions 1 and 2 in experiment 2 were not tested because Moore and Sek (2009a) showed that when the bandpass filter for the TFS1 test was centered at 11 times F0, normal-hearing subjects could only consistently perform the TFS1 test for center frequencies above 1100 Hz. The test was not used for a center frequency corresponding to that of region 5 because, at the time that the data were collected, the TFS1 test had not been validated for use with normal-hearing subjects at this frequency. More recently, Moore and Sek (2009b) showed that normal-hearing subjects could perform the task for frequencies up to at least 8000 Hz.

Complexes were presented at a level of 20 dB sensation level (SL), in threshold-equalizing noise (Moore *et al.*, 2000). The noise level, specified in a 1-ERB_N wide band centered at 1000 Hz, was 15 dB below the overall level of the complexes. The noise was designed to mask any combination tones and components on the skirts of the filter that were well away from the pass band. For the normal-hearing subjects, components as low as the seventh might have been audible in the background noise, and such components might have been partially resolved. However, the hearing-impaired subjects probably had broader auditory filters than normal, so it seems unlikely that they would have been able to resolve any components.

2. Procedure

A two-interval, two-alternative forced-choice task was used. Each interval contained four consecutive 200-ms complex tones. In one interval, all of the complex tones were harmonic (HHHH), and in the other, the tones alternated between harmonic and frequency shifted (HSHS). The subject was required to identify the interval that contained the frequency-shifted tones. The shift δ was varied adaptively in a two-up, one-down procedure. It was increased by a factor of k following an incorrect response, and reduced by a factor of k following two consecutive correct responses. For trials before the first turnpoint, k was equal to 1.953 (1.25^3). Then k was set to 1.5625 (1.25^2) until the second turnpoint had occurred; for trials after the second turnpoint, k was equal to 1.25. The threshold corresponding to 70.7% correct was estimated as the geometric mean of the values of δ at the last six turnpoints.

The largest difference between the H and S tones occurs for a frequency shift of 0.5F0. For some runs, a frequency shift larger than this was dictated by the adaptive procedure. If this happened three times in a run, then the test switched to a non-adaptive procedure. For these cases, the frequency shift was set at 0.5F0, 20 further trials were conducted, and the percentage of correct responses was measured.

3. Task familiarization and testing

The absolute detection thresholds for pure tones at each center frequency were determined at the beginning of the testing session using a two-down one-up adaptive procedure, so that the level of the H and S tones could be set accurately at 20 dB SL.

Hopkins and Moore (2007) reported that some hearing-impaired subjects performed no better than chance in a task similar to the one used here. To make sure that all subjects understood the task, a run was included that used the same procedure as the testing run, but different stimuli. Subjects were required to discriminate two harmonic tones with different F0s, rather than H and S tones. The same procedure and center frequencies were used as for testing. All of the subjects could perform the adaptive procedure for this task. Moore and Sek (2009a) reported only very small training effects for the stimuli and procedure described here. Consequently, just one training run was included with the test stimuli for each condition. Subjects then completed three test runs for each condition.

4. Analysis

In order to compare thresholds obtained from the adaptive procedure and percent correct scores from the non-adaptive procedure, both measures were converted to d' values (Green and Swets, 1974), using the method described by Hopkins and Moore (2007). Conversion was by means of a table of d' values for m-alternative forced-choice procedures (Hacker and Ratcliff, 1979). The adaptive procedure tracked the 70.7% correct point on the psychometric function, which corresponds to a d' of 0.78 for a two-alternative forced-choice task. When the adaptive procedure was used, the d' value that would have been measured for a difference of 0.5F0 Hz was calculated by dividing 0.78 by the threshold measured in the adaptive procedure, and by multiplying this value by 0.5F0. This is based on the assumption that d' is proportional to δ. The square root of the absolute value of the mean d' (with the sign of the values restored following transformation) was used in all of the statistical analyses, as this transformation results in roughly equal variance across conditions (Hopkins and Moore, 2007).

B. Results

A two-way, mixed model ANOVA was performed to assess the between-subject effect of subject type (normal-hearing or hearing impaired) and the within-subject effect of center frequency (1416 or 2904 Hz) on performance. The effect of subject type was highly significant [$F(1,12)=40.5$; $p<0.001$], but there was no significant effect of center frequency, or interaction between the factors [$F(1,12)=2.5$; $p=0.19$ and $F(1,12)=0.3$; $p=0.59$, respectively]. Consequently, further analysis was carried out on the mean d' scores for the two center frequencies for each subject. Figure 6 shows the benefit from adding TFS information to the whole frequency spectrum (bottom, quantified as the difference between SRTs for the unprocessed condition and those for the condition where all frequency channels were vocoded, as measured in experiment 2) and the d' values for

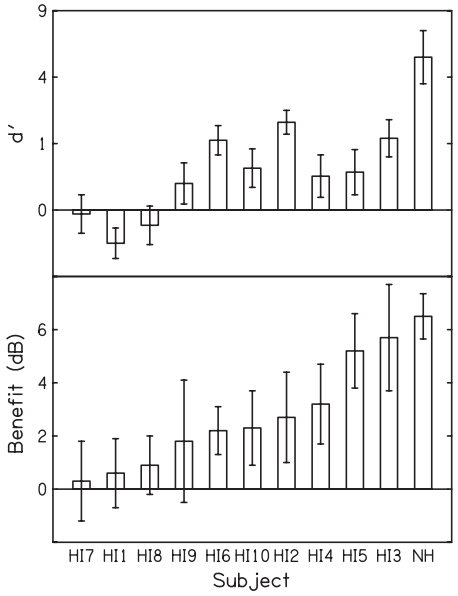


FIG. 6. Individual results for the hearing-impaired subjects and mean results for the normal-hearing subjects for experiments 2 and 3. The benefit from adding TFS information to the whole spectrum, as measured in experiment 2, is plotted in the bottom panel, ordered by performance. Scores from experiment 3 are plotted in the top panel, in the same order. Error bars show ±1 standard error of the mean.

TFS sensitivity (top), for each hearing-impaired subject. The mean results for the normal-hearing subjects are shown for comparison (note that different normal-hearing subjects took part in experiments 2 and 3, and so comparison of the individual data was not possible). The results are plotted with subjects in order of benefit from TFS as measured in experiment 2, and the results for the TFS1 test are plotted with subjects in the same order. For the hearing-impaired subjects, the benefit from adding TFS information in experiment 2 was significantly correlated with the mean TFS1 score from experiment 3 (Pearson's correlation coefficient, $r=0.66$, $p=0.04$). The same data are represented as a scatter plot in Fig. 7, with a regression line shown for the data from the hearing-impaired subjects. The line was calculated using a method described by Press *et al.* (1992), which minimizes the perpendicular distance between the points and the regression line, as is recommended for situations when there is variability in the data plotted on both axes. There was no

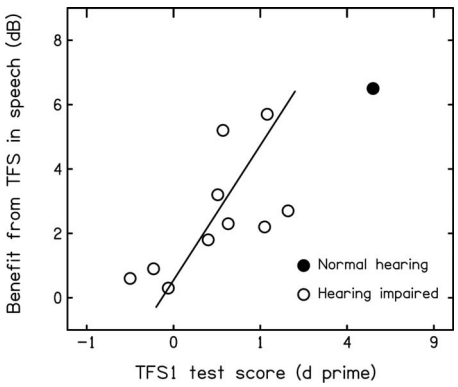


FIG. 7. As Fig. 6, but with the data plotted as a scatter plot, and a regression line fitted to the data for the individual hearing-impaired subjects.

significant correlation between the mean TFS1 scores and the mean absolute thresholds that were measured at the same center frequencies (1419 and 2904 Hz) for the hearing-impaired subjects ($r=-0.47$, $p=0.17$). However, the correlation between the means of the audiometric thresholds at 250, 500, 1000, and 2000 Hz, and the mean TFS1 scores, was significant ($r=-0.63$, $p=0.05$). The correlation between the mean audiometric threshold and the benefit from TFS information in speech was not significant ($r=-0.491$, $p=0.15$). When the effect of the mean audiometric threshold was partialled out, the correlation between the mean TFS1 test score and the benefit from adding TFS to the whole speech signal was not significant ($r=0.51$, $p=0.16$).

C. Discussion

The significantly better performance of the normal-hearing subjects than of the hearing-impaired subjects in discriminating complex tones on the basis of their TFS is consistent with the results of Hopkins and Moore (2007), who showed that hearing-impaired subjects with moderate cochlear hearing loss were very poor at discriminating similar stimuli. The results for the normal-hearing subjects are consistent with the results of Moore and Sek (2009a), who used the same method as was used here. They found that, for an F0 of 200 Hz and center frequency of 2200 Hz, normal-hearing subjects with little training had a mean threshold of 22 Hz, which is equivalent to a d' of 3.5, calculated using the method described in Sec. VI A 4.

Together, the results of experiments 2 and 3 support the idea that hearing-impaired subjects are less able to use TFS information than normal-hearing subjects, which is consistent with previous results (Buss *et al.*, 2004; Lacher-Fougère and Demany, 2005; Lorenzi *et al.*, 2006; Moore *et al.*, 2006a; Hopkins and Moore, 2007; Hopkins *et al.*, 2008). A deficit in TFS processing could occur by a number of mechanisms. A deficit in phase locking would reduce the fidelity of transmission of TFS to the central auditory system. Such a deficit was found in a physiological study of animals with induced hearing loss (Woolf *et al.*, 1981), although in a similar study such a deficit was not found (Harrison and Evans, 1979). It remains unclear whether such a deficit exists in human listeners with cochlear hearing loss. Another possibility is that the TFS processing deficits arise as a result of the broadened auditory filters that are typically found in subjects with cochlear hearing loss (Glasberg and Moore, 1986). The TFS at the output of a broader auditory filter would be more complex than for a narrower filter with the same center frequency, and this more complex information could be uninterpretable by the central auditory system.

The correlation between the benefit obtained from the addition of TFS to the whole speech signal and the sensitivity to TFS, measured using the TFS1 test, was modest, but significant. However, the correlation decreased and became non-significant when the effect of the mean audiometric threshold was partialled out. This could be taken as indicating that the correlation between the measures of sensitivity to TFS obtained in experiments 2 and 3 occurred partly because both were related to the absolute threshold. However, this

seems unlikely for the following reasons. First, scores for the TFS1 test were not significantly correlated with absolute thresholds at the center frequencies used for the TFS1 test. Second, the benefit from TFS information in speech, as measured in experiment 2, was not significantly correlated with the mean audiometric threshold.

The benefit scores for the addition of TFS information to the whole speech signal were associated with large standard errors for individual subjects. The correlation between the measures of sensitivity to TFS for experiments 2 and 3 might have been larger if the benefit scores had been measured more accurately. Also, the correlation might have been more significant if a larger number of subjects had been tested. It was hoped that the benefit of adding TFS information to speech in a given spectral region would be predicted by the sensitivity of a subject to TFS measured using the TFS1 test at the same center frequency. However, the present results were too variable to test this hypothesis: there was no significant benefit of adding TFS information to a narrow spectral region, either for the hearing-impaired subjects as a group, or for most of the individual subjects. It is not clear whether this lack of significant benefit reflected a complete inability to use TFS information in a narrow spectral region, or whether significant benefits would have been measured, had the SRTs for each condition been determined more accurately. Also, Moore and Sek (2009a) showed that when the bandpass filter was centered at 11 times F0, normal-hearing subjects could only consistently perform the TFS1 test for center frequencies above 1100 Hz. Consequently, sensitivity to TFS could only be assessed using the TFS1 test for center frequencies corresponding to two of the five spectral regions that were tested. The TFS1 test cannot be used to assess sensitivity to TFS information for frequencies below 1000 Hz, where TFS information is expected to be important in coding F0 information.

VII. CONCLUSIONS

For normal-hearing subjects, the largest benefit (reduction in SRT) of adding TFS information to a vocoded target and competing-talker signal was obtained when TFS information was added to channels with center frequencies below 1000 Hz. This was the case regardless of whether TFS information was added cumulatively, starting from the low-frequency or the high-frequency channels. This is consistent with the idea that TFS is important for coding F0 information. However, some benefit from adding TFS information was measured when TFS was added to the high-frequency channels only. A corresponding benefit was not measured when TFS was added to the same channels when TFS was already present in the low-frequency channels, suggesting that there is some redundancy in TFS information across frequencies.

In a second experiment, the improvement in SRT, when TFS information was added to each of five spectral regions (spanning 100–8000 Hz) was similar across regions for normal-hearing subjects, suggesting that TFS is important for coding information over and above that related to F0.

Hearing-impaired subjects as a group did not show a significantly greater benefit when TFS information was added to an isolated spectral region than when envelope information alone was added. The benefit from adding TFS information to the whole frequency spectrum was correlated with a psychophysical measure of TFS sensitivity across individual listeners, which adds support to the idea that sensitivity to TFS information is important for listening to speech in a complex background.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Michael Stone and Brian Glasberg for their contributions to this work. We also thank Hedwig Gockel, Joshua Bernstein, an anonymous reviewer, and the Associate Editor Richard Freyman for helpful comments and suggestions.

Bernstein, J. G., and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 3358–3372.

Bernstein, J. G., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

Bracewell, R. N. (**1986**). *The Fourier Transform and Its Applications* (McGraw Hill, New York).

Bregman, A. S., Liao, C., and Levitan, R. (**1990**). "Auditory grouping based on fundamental frequency and formant peak frequency," Can. J. Psychol. **44**, 400–413.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., and Nelson, D. A. (**1986**). "Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **79**, 1034–1044.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

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

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.

Harrison, R. V., and Evans, E. F. (**1979**). "Cochlear fibre responses in guinea pigs with well defined cochlear lesions," Scand. Audiol. Suppl. **9**, 83–92.

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.

Hilbert, D. (**1912**). *Grundzuege einer Allgemeinen Theorie der Linearen Integralgleichungen (Basic Principles of a General Theory of Linear Integral Equations)* (Teubner, Leipzig).

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., and Moore, B. C. J. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J.

Acoust. Soc. Am. **125**, 442–446.
Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "The effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.
Kitterick, P. T., and Summerfield, A. Q. (**2007**). "The role of attention in the spatial perception of speech," in Proceedings of the MidWinter Meeting of the Association for Research in Otolaryngology, Denver, CO.
Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.
Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C. J., and Glasberg, B. R. (**1998**). "Use of a loudness model for hearing aid fitting. I. Linear hearing aids," Br. J. Audiol. **32**, 317–335.
Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006b**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.
Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006a**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.
Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). "New version of the TEN test with calibrations in dB HL," Ear Hear. **25**, 478–487.
Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.
Moore, B. C. J., and Sek, A. (**2009a**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161–171.
Moore, B. C. J., and Sek, A. (**2009b**). "Sensitivity of the human auditory system to temporal fine structure at high frequencies," J. Acoust. Soc. Am. **125**, 3186–3193.
Moore, G. A. and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.
Moore, T. (**1981**). "Voice communication jamming research," in Proceedings of the AGARD Conference, 331: Aural Communication in Aviation, Neuilly-Sur-Seine, France.
Oxenham, A. J., Micheyl, C., and Keebler, M. V. (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?," J. Acoust. Soc. Am. **125**, 2189–2199.
Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., (**1992**). *Numerical Recipes in C* (Cambridge University Press, Cambridge, England).
Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470–479.
Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.
Souza, P., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.
Steeneken, H. J. M., and Houtgast, T. (**1999**). "Mutual dependence of the octave-band weights in predicting speech intelligibility," Speech Commun. **28**, 109–123.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Stone, M. A., Fullgrabe, 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.
Summers, V., and Leek, M. R. (**1998**). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," J. Speech Lang. Hear. Res. **41**, 1294–1306.
Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. (**1998**). "Frequency-weighting functions for broadband speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580–1585.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Heifer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of outer hair cells," Hear. Res. **4**, 335–346.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibres," J. Acoust. Soc. Am. **66**, 1381–1403.
Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues (L)," J. Acoust. Soc. Am. **116**, 1351–1354.
Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise

David. L. Woods[a)]
Department of Neurology, UC Davis and VANCHCS, 150 Muir Road, Martinez, California 95553

E. William Yund, Timothy J. Herron, and Matthew A. I. Ua Cruadhlaoich
Research Service, VANCHCS, 150 Muir Road, Martinez, California 95553

(Received 20 March 2009; revised 16 December 2009; accepted 22 December 2009)

Identification functions of 20 initial and 20 final consonants were characterized in 9600 randomly sampled consonant-vowel-consonant (CVC) tokens presented in speech-spectrum noise. Because of differences in the response criteria for different consonants, signal detection measures were used to quantify identifiability. Consonant-specific baseline signal-to-noise ratios (SNRs) were adjusted to produce a d' of 2.20 for each consonant. Consonant identification was measured at baseline SNRs (B), at B−6, and at B+6 dB. Baseline SNRs varied by more than 40 dB for different consonants. Confusion analysis revealed that single-feature place-of-articulation errors predominated at the highest SNR, while combined-feature errors predominated at the lowest SNR. Most consonants were identified at lower SNRs in initial than final syllable position. Vowel nuclei (/ɑ/, /i/, or /u/) significantly influenced the identifiability of 85% of consonants, with consistent vowel effects seen for consonant classes defined by manner, voicing, and place. Manner and voicing of initial and final consonants were processed independently, but place cues interacted: initial and final consonants differing in place of articulation were identified more accurately than those sharing the same place. Consonant identification in CVCs reveals contextual complexities in consonant processing.
[DOI: 10.1121/1.3293005]

PACS number(s): 43.71.Es, 43.71.Sy, 43.71.An [MSS] Pages: 1609–1623

I. INTRODUCTION

Although most spoken syllables contain multiple consonants, consonant identification of large consonants sets in multi-consonant syllables has received relatively little study. The current experiment investigated the identification of 20 initial and 20 final consonants in a pseudorandomly sampled set of 9600 consonant-vowel-consonant (CVC) tokens presented in speech-spectrum noise.

A. Quantifying consonant identification

In their classic study, Miller and Nicely (1955) found that the identifiability of different consonants presented in noise varies substantially: some consonants (e.g., nonsibilant fricatives) were difficult to identify even at high signal-to-noise ratios (SNRs), while others (e.g., sibilants) were accurately identified at much lower SNRs. For example, /ʃ/ hit rates at SNRs of −12 dB were similar to /ð/ hit rates at SNRs of +6 dB. Their results also suggested that comparisons of consonant identifiability using hit rate alone are potentially confounded by response biases associated with different consonants. For example, they found that hit rates for /v/ were much higher than hit rates for /ð/. However, at low and intermediate SNRs, the majority of /v/ responses were elicited by other consonants, particularly /ð/. Indeed, CVs containing /ð/ elicited more false /v/ responses than /ð/ hits. Similarly, Wang and Bilger (1973) found that CVs containing /ð/ elicited nearly twice as many false /v/ responses as /ð/ hits, while false /ð/ responses were rarely elicited by CVs containing /v/. Such asymmetric confusion patterns suggest that subjects have significant response biases: among confusable consonants, ambiguous phonological cues are much more likely to elicit the report of some consonants (e.g., /v/) rather than others (e.g., /ð/).

As a result, hit rate alone can misrepresent consonant identification accuracy. Metrics are needed that incorporate information from both hit and false response rates. Signal detection theory (Green and Swets, 1974) provides two such metrics: d', a measure of receiver sensitivity that quantifies the discriminability of sensory information, and beta, a criterion metric that quantifies the amount of sensory evidence needed to elicit a particular response. Therefore, the current experiments incorporated signal detection measures to quantify consonant identification performance following similar methods used in previous studies (Müsch and Buus, 2001b, 2001a).

B. Consonant identifiability in noise

Several previous studies have investigated the identification of small consonant sets in CVC syllables. Boothroyd and Nittrouer (1988) developed matched sets of 120 words and 120 nonsense syllables by combining ten initial consonants, ten vowels, and ten final consonants. Although they did not analyze the identification of individual consonants, their results suggested that initial consonants, final consonants, and vowels were processed independently in nonsense syllables. Benkí (2003) presented Boothroyd and Nittrouer's

[a)]Author to whom correspondence should be addressed. Electronic mail: dlwoods@ucdavis.edu

(1988) nonsense CVC tokens to 37 undergraduates and analyzed consonant-confusion patterns at four different SNRs spanning a range of 9 dB. Because of the small number of CVC tokens, separate subjects were tested at each SNR to avoid token repetition. Benkí (2003) found large differences in the identifiability of different consonants. For example, the sibilant /s/ was almost perfectly identified even at the most difficult SNR, whereas the non-sibilant fricative /f/ was frequently misidentified even at the highest SNR. These results imply that different consonants require markedly different SNRs for identification. However, because Benkí (2003) used only ten consonants, confusions between many potentially confusable consonant pairs (e.g., /s/ and /ʃ/) could not be examined. Moreover, because only 240 of 1000 possible CVC combinations were used, it was impossible to independently analyze the processing of initial consonants, final consonants, and vowels. Redford and Diehl (1999) used a completely randomized set of CVCs created by using the consonants /p/, /t/, /k/, /f/, /θ/, /s/, and /ʃ/) at initial and final positions combined with the three vowels /ɑ/, /i/, and /u/. The CVCs were presented at mid-sentence in different sentence contexts at SNRs of +15 dB. Significant differences were again found in consonant identification accuracy with non-sibilant fricatives proving particularly difficult to identify (e.g., /f/ and /θ/).

Previous studies of consonant identification in CVCs used small consonant sets and thus may have overestimated the identifiability of many consonants. For example, the unvoiced plosives (/p/, /t/, and /k/) are often confused with each other, but may also be confused with voiced plosives, affricates, /h/, and other unvoiced fricatives. The study of consonant identification in larger consonant sets has been largely restricted to single consonant (CV and VC) syllables. Miller and Nicely (1955) analyzed the identification of 16 common initial consonants in CVs with the vowel /ɑ/. There were five female participants, four of whom identified consonants in the CVs spoken by the fifth. Talkers rotated on successive blocks so that a total of approximately 7000 different tokens were presented at seven different SNRs ranging from −18 to +12 dB. Confusion matrices for 4000 CV token presentations (approximately 250 presentations of each consonant) were obtained at each SNR and revealed large differences in the identifiabilities of different consonants. For example, nasals were more accurately identified at −12 dB than non-sibilant fricatives (e.g., /θ/) at +6 dB. The percentage of consonants correctly identified improved with SNR, with mean hit rates increasing from 7.8% at −18 dB to 90.6% at +12 dB, producing an average performance/SNR (P/S) slope of 2.8%/dB. They found that the slopes of P/S functions were generally steepest around SNRs near mid-performance (e.g., 50% correct) levels. However, differences in the P/S slopes were evident for different consonants. For example, plosives had steeper P/S slopes (e.g., /t/ 5.9%/dB) than non-sibilant fricatives (e.g., /θ/ 2.3%/dB).

Wang and Bilger (1973) studied consonant identification in broadband noise using two different CV and VC lists, each with 16 consonants. They found large differences in consonant identification for both CVs and VCs, with non-sibilant fricatives again proving particularly difficult to identify. Phatak and Allen studied consonant identification in both speech-spectrum noise (Phatak and Allen, 2007) and white noise (Phatak *et al.*, 2008) using CVs containing 16 consonants paired with four vowels. Nasal consonants proved easy to identify in white noise but hard to identify in speech-spectrum noise. Again, non-sibilant fricatives proved particularly difficult to identify in both noise conditions. For example, /ð/ hit rates in speech-spectrum noise at 0 dB SNR were similar to /z/ hit rates at −22 dB SNR.

C. The identification of leading and trailing consonants

Several previous studies of CVCs have reported that initial consonants are more accurately identified than final consonants. For example, Benkí (2003) found lower error rates for initial than final consonants at all SNRs tested. Redford and Diehl (1999) also found that initial consonants were more accurately recognized than final consonants across most listening conditions. However, Redford and Diehl (1999) noted that the initial consonant advantage was absent for some consonants (e.g., /θ/) and was influenced by vowel context, being enhanced in syllables containing /ɑ/. Similar initial consonant advantages have been found in comparisons of CVs and VCs (Wang and Bilger, 1973; Dubno and Levitt, 1981). However, a recent study investigating consonant identification in CVs and VCs found the opposite result: final consonants were identified more accurately than initial consonants in six-talker speech babble (Cutler *et al.*, 2004). These unexpected results may have reflected differences in initial and final consonant masking levels. The babble-noise masker was created by manually selecting 1 s segments of continuous speech from each of six different talkers and then combining the six speech segments. This procedure may have resulted in a better alignment of the speech-babble amplitude envelopes at babble onset so that greater masking would have occurred for initial than final consonants.

D. Analyzing phonetic features

Consonant identification requires the accurate analysis of place, manner, and voicing features. Considerable evidence suggests that these features differ in their discriminability in noise. For example, Dubno and Levitt (1981) studied consonant confusions in speech-spectrum noise in 11 different CV and VC lists. In each list, seven, eight, or nine different consonants were presented with a single vowel (/ɑ/, /i/, or /u/) in speech-spectrum noise. They found higher rates for place than for manner errors, and higher rates for place errors than combined place+manner (P+M) errors. Because voiced and unvoiced consonants were presented in separate lists, voicing errors could not be evaluated. Other studies using larger consonant sets including both voiced and unvoiced consonants also found that place errors were more common than manner errors, and that both place and manner errors were more common than voicing errors (Miller and Nicely, 1955; Wang and Bilger, 1973; Levitt and Resnick, 1978; Helfer and Wilber, 1990; Gelfand *et al.*, 1992; Phatak and Allen, 2007; Phatak *et al.*, 2008).

E. Vowel influences on consonant identifiability

Previous studies also found that vowels influence consonant identifiability. For example, Wang and Bilger (1973) found that consonant identification varied with vowel: consonant report was most accurate for syllables containing /ɑ/, intermediate for syllables with /i/, and lowest for syllables containing /u/. Moreover, vowel context affected the identifiability of leading and trailing consonants differentially: leading consonants were more accurately identified in syllables containing /ɑ/ and trailing consonants were more accurately identified in syllables containing /i/. Dubno and Levitt (1981) also found an overall superiority of consonant processing in syllables containing /ɑ/. In particular, nasals were identified approximately twice as accurately in syllables containing /ɑ/ than in syllables containing /i/ or /u/. In contrast, affricates were most accurately reported in syllables containing /i/. In addition, they found an interaction between place of articulation and vowel: front and middle consonants were less accurately reported in syllables containing /i/ than /u/, whereas the reverse pattern was found for back consonants. Finally, in CVCs, Redford and Diehl (1999) found that vowel effects differed for initial and final consonants within a syllable: initial consonants were most accurately reported in syllables containing /ɑ/, whereas final consonants were most accurately reported in syllables containing /u/.

F. Interactions in identifying initial and final consonants

Interactions in the processing of initial and final consonants in CVCs have not been systematically examined. However, consonant-consonant interactions are known to alter the processing of place-of-articulation features over syllable-length silent gaps (Mann and Repp, 1981). This raises the possibility that interactions in the processing of initial and final consonant features might be observed in CVCs, for example, if stimulus-specific neuronal adaptation (Woods and Elmasian, 1983) reduced the salience of repeated consonant features.

G. The current experiment

The goal of the current experiment was to characterize the identification of consonants in CVCs that included 20 initial and final consonants combined with three vowels (/ɑ/, /i/, and /u/). In order to estimate average consonant confusions, performance measures were obtained from the pseudorandom sampling of 9600 different CVC syllable tokens. The large consonant set permitted the examination of an extensive set of potential consonant confusions. Signal detection theory was used to measure identification performance for each consonant over a consonant-specific range of SNRs adjusted to equate overall identification performance for different consonants. This permitted the characterization of both baseline identifiability and P/S functions for each consonant in initial and final syllable positions. This further allowed comparisons of the identifiabilities of consonant groups sharing similar place, manner, or voicing features as well as an analysis of the patterns of feature-processing errors (single-feature place, manner, voicing errors, and combined-feature errors) in initial and final syllable positions at different SNRs. Finally, the use of CVCs permitted an analysis of the effects of the vowel nucleus on the identifiability of initial and final consonants as well as the evaluation of possible interactions in the processing of initial and final consonants within a CVC.

II. METHODS

The present report describes consonant identification and consonant confusions observed in 34 560 trials, obtained from 16 young normal-hearing subjects who each participated in three separate 1 h testing sessions. Additional details about test-retest reliability, the relationship of consonant identification performance to audiometric thresholds, interactions in the processing of initial and final consonants in word and nonsense syllable tokens, and the relationship of consonant identification performance to sentence reception thresholds are provided elsewhere (Woods *et al.*, 2010).

A. Subjects

Sixteen young subjects (eight female and eight male, ages 18–30 yrs) with normal hearing (thresholds ≤20 dB hearing level at 250–4000 Hz)[1] each participated in three testing sessions within 11 days.

B. Syllable tokens

The CVC list included 1200 syllables constructed from the exhaustive combination of 20 initial consonants, 20 final consonants, and three vowels (/ɑ/, /i/, and /u/). Twenty-one consonants (/b/, /d/, /g/, /r/, /l/, /ŋ/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, /k/, and /h/) were used, 19 in both initial and final syllable positions, /h/ occurring in only the initial position, and /ŋ/ in only the final position. The token corpus was created by first recording 4800 syllable sets from each of four phonetically trained talkers (two male and two female) using an AKG C-410 head mounted microphone in an Industrial Acoustics Company (New York, NY) sound booth. The four talkers had been raised in different parts of the United States (two in the Midwest and two in California) and had slightly different American English speech patterns. Syllables were digitized at 16 bit resolution and 44.1 kHz sampling rate using MATLAB (The MathWorks, Inc., Natick, Ma). The complete syllable sets were reviewed by one of the authors (E.W.Y.) and the two best exemplars of each syllable (2400 tokens) were selected from each talker's corpus. Then, two listeners with normal hearing independently reviewed each of the 9600 syllables in the absence of masking noise to assure the intelligibility of all tokens. Whenever the intelligibility test failed, a new exemplar was substituted and further testing was performed among laboratory staff to assure the intelligibility of the substituted tokens. Within the entire 9600 token corpus syllable durations ranged from 350 to 890 ms (mean: 636 ms). For each token, 100 ms at the center of each vowel was identified by manual review to establish the time interval over which the noise masking levels would be linearly adjusted (on the dB scale) to provide appropriate masking levels for initial and final consonants differing in intrinsic identifiability (Fig. 1).

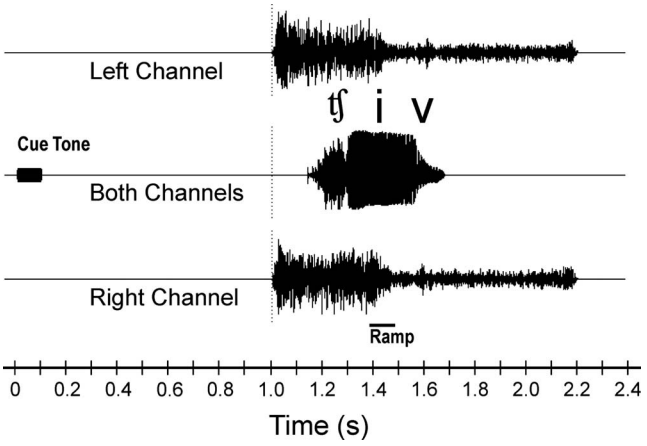


FIG. 1. Trial structure. Trials were cued by a 1.0 kHz tone. After 1.0 s two independent 1200 ms noise bursts were presented through the left and right loudspeakers. CVCs were presented simultaneously through both loudspeakers at random intervals after noise-burst onset. Noise amplitudes were modulated over a 100 ms interval during the mid-vowel segment of the CVC to provide appropriate masking levels for different initial and final consonants.

C. Speech-spectrum noise adjustment

Talker-specific speech-spectrum noise was used to mask each CVC token. The average spectrum for each talker was first obtained by averaging the spectra of all CVC tokens for that talker. This spectrum was then used to create a finite impulse response filter that was used to filter broadband white noise. Each filtered-noise file was trimmed of the first 0.5 s and cut into 100 different noise segments of 1200 ms duration each, which were then randomly sampled during the testing sessions.

D. Stimuli and procedure

Testing was performed in a 2.44×2.44 m single-walled, sound-attenuating testing room. The interior walls were covered by 2.5 cm thick acoustic foam resulting in ambient third-octave noise levels less than 20 dB sound pressure level (SPL) from 250 to 4000 Hz. In anticipation of future studies of hearing-impaired subjects with hearing aids, stimuli were presented through loudspeakers (M-Audio Studiophile AV 40). Immediately before the first CVC session, subjects were briefed with written and oral instructions and received ~5 min of training in identifying CVCs presented without masking noise.

During each session, each subject was presented with 720 different CVCs selected by constrained random sampling of the 9600 tokens. The CVCs were grouped by talker into 30-trial blocks with 24 talker-specific blocks presented during each experiment. Each trial (Fig. 1) began with a tone burst cue (100 ms, 1.0 kHz tone, 70 dB SPL). After a delay of 1.0 s, two randomly selected talker-specific noise bursts were presented independently through the left and right speakers, along with a single CVC presented synchronously from both loudspeakers. CVC onset times were randomized with respect to noise-burst onset with the constraint that each CVC began at least 100 ms after the onset of the noise bursts and ended at least 100 ms before their offset.

On each trial, the listener verbally reported the CVC token, referring, when necessary, to a list of acceptable initial and final consonants and vowels posted in the testing room. Responses were spoken into a microphone and phonetically transcribed using a modified QWERTY keyboard by an investigator (M.A.I.U.C.) listening through headphones in an adjacent room. Responses were spoken in quiet and subjects were queried via intercom when responses were invalid or poorly enunciated.[2] Subjects were given the option of repeating trials in cases of attentional lapse or noise interference (e.g., coughing, etc.). Each inter-trial interval of approximately 2 s included the time needed for syllable transcription plus a small delay (0.5 s) before the delivery of the warning tone signaling the next trial. Trials occurred at a rate of approximately 15/min so that each 720-syllable test required about 48 min in addition to rest breaks that occurred at each subject's discretion. Presentation software (NEUROBEHAVIORAL SYSTEMS, version 12.0) was used for stimulus delivery, masking noise and signal level adjustment, response monitoring, and d' calculations.

Syllable intensity was randomly roved from 70 to 75 dB SPL in 1 dB steps. Psychometric functions were measured for each initial and final consonant at three different SNRs: B−6, B, and B+6 dB relative to a baseline SNR (B) that was specific to each initial and final consonant. Consonant-specific SNR levels were established in preliminary experiments (see below). The SNR level (i.e., B−6 dB, B, or B +6 dB) varied randomly across trials.

During each test session, 720 tokens were randomly selected without repetition from the syllable corpus of 9600 tokens. Selection was constrained so that each initial and final consonant was presented exactly 12 times at each SNR. The 12 tokens included four syllables containing each of the three vowels /ɑ/, /i/, and /u/, with each syllable spoken by a different talker. CVC selection was selected based on the initial consonant, vowel, and final consonant so that each consonant was presented at initial and final syllable positions on 36 trials (i.e., 12 at each SNR) and each vowel was presented on 240 trials (80 at each SNR). Each token in the corpus had an equal probability of being presented. Following talker and CVC selection, one of the two CVC tokens spoken by that talker was randomly selected. This procedure resulted in the presentation of 240 tokens (60 from each talker) at each of the three baseline SNR levels (B−6, B, and B+6 dB). Because of the limited set of vowels and the low rate of vowel errors, only consonant identification was scored.

E. Baseline SNR adjustments to equate consonant identifiabilities

Two preliminary experiments were used to estimate the baseline SNRs needed to equate the identifiabilities of different consonants. In the first experiment, four members of the laboratory staff were tested repeatedly (22, 1 h sessions) while consonant-specific SNR levels were adjusted independently in initial and final syllable positions to equate the hit rates for different consonants. This approach revealed significant differences in the response criteria for different consonants. Therefore, signal detection theory was used to com-

TABLE I. SNRs used for the presentation of initial and final consonants in the main experiment.

	b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	10.3	3.8	6.9	2.5	10.5	⋯	5.7	7.8	10.6	19.6	−2.6	−0.4	0.9	1.1	−7.4	19.7	9.1	12.7	−1.5	6.9	15.5
Final	13.0	11.6	13.8	−0.3	4.4	19.3	17.5	17.5	21.4	32.6	−1.8	4.5	0.4	0.3	−7.9	17.6	11.1	12.3	3.2	7.5	⋯

pute d' for each consonant using hit and false alarm (FA) rates. FAs were defined as consonant responses that occurred when a different consonant had been presented (e.g., a /v/ response when /ð/ was actually presented). An examination of the confusion matrices suggested that FAs did not occur randomly to all possible consonants, but rather occurred in response to a small group of potentially confusable consonants. For example, /p/ FAs occurred in response to /t/, /k/, /h/, /b/, and /g/ but were rarely observed in response to other consonants. On average, significant FAs (exceeding approximately 1% of total responses including hits) were produced by approximately 6.5 other consonants. Therefore, the FA rate was calculated by assuming a pool of 6.5 FA-producing stimuli. While assuming that a fixed pool size of FA-producing consonants was somewhat arbitrary, it eliminated the problem of estimating the FA-pool size for consonants identified using different criterion levels (see discussion below).

The results of the first preliminary study were used to adjust SNR levels to produce an estimated mean d' of 2.6 for each consonant that were used in the second preliminary experiment. Seven young naive volunteers with normal hearing participated after giving informed consent following local Institutional Review Board regulations. Each subject was tested on three occasions for a total of 21, 1 h sessions. Calculated d' values in the second preliminary experiment averaged 2.52 (72.9% correct) for initial consonants and 2.58 (74.8% correct) for final consonants.

The results of the second preliminary experiment were then used to make finer B-level adjustments in order to minimize variations in the identifiabilities of different consonants in the main experiment. In addition, the target d' was reduced to 2.20 (approximately 65% correct) to measure steeper portions of P/S functions. Additional baseline adjustments were also incorporated to equate performance for syllables spoken by different talkers (syllables spoken by female talkers were reduced by 1.8 dB on average) and for syllables containing different vowels (syllables with /i/ were reduced by 3.0 dB and those containing /ɑ/ were reduced by 1.2 dB relative to those containing /u/). The mean SNR values used for each consonant in the main experiment are shown in Table I. B values averaged 6.6 dB for initial consonants and 9.9 dB for independently adjusted final consonants. The range of B values needed to equate identifiabilities spanned 27.1 dB for initial consonants and 40.5 dB for final consonants.

F. Statistical analysis

The data were analyzed with analysis of variance (ANOVA) for repeated measures using the open-source CLEAVE program (T. J. Herron). The original degrees of freedom are reported for each test with the significance levels adjusted using the Box/Greenhouse–Geisser correction for inhomogeneity of variance when appropriate (Greenhouse and Geisser, 1959). In these cases, the original degrees of freedom are reported along with corrected significance (p) levels. Because of the large number of ANOVA comparisons, a relatively strict criterion ($p<0.01$) was used to evaluate statistical significance.

III. RESULTS

A. Quantifying consonant identification

On average, subjects correctly identified 64.1% of initial consonants and 65.2% of final consonants. However, although SNRs had been adjusted to equate d' values for all consonants, hit rates varied substantially for both initial consonants (range: 43.5%–82.5%) and final consonants (range: 43.2%–85.0%). Differences in hit rate were frequently accompanied by corresponding differences in FA rate. For example, the hit rate for /v/ was nearly twice as high as the hit rate for /s/ while the FA rate for /v/ was 8.8 times greater than the FA rate for /s/. Thus, these results revealed systematic differences in consonant response criteria similar to those observed in previous studies (Miller and Nicely, 1955; Wang and Bilger, 1973; Dubno and Levitt, 1981). The covariation in hit and FA rates was the primary reason that signal detection theory was used to analyze the results.

B. Consonant identifiability in noise

Analysis was first performed on initial and final consonants, averaged over vowels, speakers, and final or initial consonants, respectively. Thus, the d' value for each initial consonant reflected 1728 responses from the 16 subjects to 480 different randomly sampled CVC tokens beginning with

TABLE II. Observed d' and beta values for each consonant in initial and final positions.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	d'	2.08	2.15	2.10	2.14	2.12		2.18	2.23	1.95	2.11	2.10	1.88	2.24	2.33	2.12	2.44	2.32	2.37	2.03	2.36	2.26
	beta	1.72	1.72	1.47	2.15	0.47		2.23	1.82	0.26	2.10	1.27	1.23	1.00	2.90	2.67	1.40	0.87	1.31	2.05	1.36	0.00
Final	d'	2.03	2.14	1.98	2.35	2.29	2.11	2.28	2.41	2.15	1.90	2.17	2.12	2.14	2.18	2.24	2.36	2.20	2.20	2.13	2.15	
	beta	1.27	1.47	1.15	2.59	2.08	2.10	0.85	1.05	−0.07	1.50	2.12	0.91	1.62	1.91	2.88	0.79	1.30	1.64	1.25	1.19	

TABLE III. Estimated SNRs (in dB) needed to produce comparable identifiability ($d'=2.20$) of all consonants in the main experiment. SEM=standard error of the mean.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	Mean	11.1	4.0	7.3	2.8	11.1		5.8	7.6	12.7	21.9	−2.0	1.1	0.6	0.2	−6.9	13.7	7.8	11.6	−0.8	6.2	15.1
	SEM	0.4	0.4	0.3	0.4	0.4		0.3	0.4	0.5	0.7	0.3	0.4	0.3	0.5	0.6	0.6	0.5	0.5	0.4	0.4	0.4
Final	Mean	14.1	11.9	15.2	−1.4	3.4	19.8	16.8	15.6	22.0	38.6	−1.7	4.9	0.7	0.4	−8.4	15.8	11.1	12.3	3.6	7.8	
	SEM	0.4	0.5	0.5	0.4	0.4	0.5	0.3	0.3	0.6	0.7	0.5	0.4	0.4	0.5	0.5	0.7	0.5	0.4	0.3	0.2	

that consonant and including 576 tokens presented at each SNR (B−6, B, and B+6). Table II shows d' and beta values for each consonant. The observed d' values averaged 2.18 for initial consonants (range 0.37d' units) and 2.19 for final consonants (range 0.48d' units). These values were similar to the targeted performance level ($d'=2.20$).

A repeated measures ANOVA was used to analyze the d' values for the 19 consonants that occurred in both the initial and final syllable positions. This analysis revealed that the mean SNRs of initial and final consonants had been effectively equated by the consonant- and position-specific SNR adjustments made on the basis of preliminary studies (position effect $[F(1,15)=0.45]$). However, the SNR adjustments had not fully corrected for small residual differences in the identifiabilities of different consonants, as reflected in a significant main effect of consonant $[F(18,270)=6.76,\ p<0.0001]$ reflecting differences in mean d' values for different consonants that ranged from 1.96 (/ð/) to 2.38 (/θ/).

The B levels needed to equate consonant identifiability at the $d'=2.20$ level based on the three-point psychometric functions of each consonant are shown in Table III. These values differed from the B values actually used (Table I) by an average of 0.09 dB (range −6.0 to +6.0 dB, excluding the final /ð/ and initial /θ/, the range was −1.9 to +2.3 dB). Standard errors of the SNR means across subjects averaged 0.62 dB. This suggests that the estimated B values needed to equate the identifiability of different consonants shown in Table III were generally precise to within less than 2.0 dB. Estimated B values spanned a range of 28.8 dB for initial consonants and 47.0 dB for final consonants.

B values varied with manner of articulation, with sibilants and affricates requiring the lowest SNRs, followed at increasing SNRs by liquids, plosives, nasals, and non-sibilant fricatives. In general, unvoiced consonants were identified at lower SNRs than voiced consonants. However, B values varied considerably even among consonant sharing similar manner and voicing. For example, among unvoiced plosives, /t/ was identified at SNRs about 10 dB lower than /p/, as was previously noted by Phatak and Allen (2007).

Beta values are also presented in Table II. There were no significant correlations between d' and beta for initial $[r(19)=0.04]$ or final $[r(19)=0.11]$ consonants. ANOVA for repeated measures was used to analyze beta values for the 19 consonants. There was a significant effect of consonant $[F(18,270)=20.79,\ p<0.0001]$ that reflected large differences in response criteria of different consonants. For example, subjects produced more than eight times more /v/ than /ʃ/ false alarms, reflected in respective beta values of 0.09 and 2.40.

Table IV shows the psychometric functions of d' and beta for the different consonants. Both beta and d' increased with increasing SNRs, but there was an insignificant correlation between their rates of increase ($r=0.26$). A repeated measures ANOVA was used to analyze the slopes of the d' P/S for the 19 consonants that occurred in both initial and final positions with consonant and position as factors. The position factor failed to reach significance. However, there was a highly significant main effect of consonant $[F(18,270)=23.68,\ p<0.0001]$. Non-sibilant fricatives (/v/, /ð/, /f/, and /θ/) had P/S functions with substantially shallower slopes (mean 0.08d'/dB) than did other consonant classes including plosives (0.15 d′/dB), sibilant fricatives (0.16d'/dB), nasals (0.16d'/dB), liquids (0.17d'/dB), and affricates (0.19d'/dB). Some of the differences in slope reflected the fact that certain consonants (e.g., the non-sibilant fricatives) remained hard to identify even at the highest SNRs. This was reflected in a highly significant correlation between the B values of each consonant and the slope of its P/S function $[r=-0.70,\ t(18)=5.82,\ p<0.001]$.

C. The identification of leading and trailing consonants

In agreement with previous reports (Wang and Bilger, 1973; Dubno and Levitt, 1981; Redford and Diehl, 1999), initial consonants were collectively detected at 3.3 dB lower SNRs than final consonants. However, this initial consonant advantage was not observed for all consonants: liquids were

TABLE IV. Slopes of psychometric functions of d' and beta for each consonant in initial and final position.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	d'/dB	0.16	0.24	0.24	0.22	0.13		0.16	0.16	0.12	0.04	0.18	0.21	0.13	0.14	0.15	0.04	0.09	0.15	0.23	0.22	0.15
	Beta/dB	0.11	0.08	0.06	0.18	0.02		0.07	0.09	0.00	0.06	−0.03	0.07	−0.02	0.09	0.14	0.02	0.00	0.04	0.15	0.01	0.07
Final	d'/dB	0.16	0.19	0.16	0.14	0.09	0.18	0.11	0.11	0.09	0.05	0.20	0.20	0.19	0.19	0.08	0.09	0.17	0.17	0.17	0.19	
	Beta/dB	0.05	0.03	0.07	0.08	0.08	0.07	−0.01	0.01	−0.01	0.09	0.03	−0.06	0.05	0.17	0.05	0.06	0.11	0.09	0.09	0.08	

TABLE V. Confusion matrices for initial consonants averaged across subjects, SNRs, and voices. Each consonant was delivered on 1728 trials.

	b	d	g	r	l	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
b	1001	12	24	17	50	2	32	266	16	4	4	2	0	5	15	69	58	7	16	128
d	32	1059	110	28	73	26	28	74	36	27	32	6	4	13	18	15	18	32	30	67
g	36	76	1097	26	62	10	26	78	20	16	22	2	3	5	11	17	41	6	33	141
r	39	35	85	924	206	27	62	114	17	29	20	16	2	12	4	15	27	12	20	62
l	28	9	11	50	1334	19	48	150	28	2	1	1	3	2	0	6	10	1	1	24
n	17	28	33	45	303	917	170	77	18	11	12	3	0	4	7	5	8	4	10	56
m	39	5	18	43	203	106	1073	127	9	6	3	1	1	2	4	7	13	1	10	57
v	57	10	12	37	61	6	15	1360	102	11	0	1	0	0	11	17	8	2	3	15
ð	17	36	6	1	193	2	1	499	861	41	1	1	0	2	54	8	0	2	0	3
z	14	51	26	12	38	10	20	54	30	1155	66	21	7	53	27	27	16	32	25	44
ʤ	12	85	78	24	53	9	23	23	9	45	1056	138	15	7	9	12	12	24	49	45
ʧ	3	7	13	2	7	2	1	4	2	15	253	1222	67	6	3	5	15	42	43	16
ʃ	5	15	9	3	12	2	4	3	1	23	278	451	813	23	5	4	10	12	17	38
s	11	24	15	10	24	6	14	39	10	424	47	52	21	734	42	64	25	58	39	69
θ	1	3	0	0	0	0	0	15	9	3	0	0	0	7	1224	455	3	6	0	2
f	25	2	0	0	4	2	2	120	8	2	1	3	0	5	196	1333	9	1	8	7
p	9	3	7	5	6	1	7	7	0	0	1	0	0	0	1	15	1264	16	58	328
t	22	47	27	10	42	8	24	34	10	39	36	41	7	21	23	25	92	880	165	175
k	12	9	30	3	7	4	7	12	2	9	6	1	1	4	8	26	116	41	1253	177
h	11	0	4	0	7	0	4	13	2	0	2	2	0	1	3	20	176	7	51	1425

detected at lower SNRs in final consonant position. There also remained small position × consonant interactions for d' [$F(18,270)=3.82$, $p<0.001$] after consonant- and position-specific B adjustments. For some consonants, d' values were higher in initial than final syllable position (e.g., /p/, /k/, and /ʃ/) but the opposite pattern was seen for other consonants (e.g., /ð/, /v/, and /s/). Beta values varied significantly with position [$F(1,15)=25.93$, $p<0.0001$] due to stricter criteria for initial than for final consonants (1.58 vs 1.46). Beta measures also showed a significant position × consonant interaction [$F(18,270)=15.89$, $p<0.0001$], reflecting the fact that some consonants showed different response criteria in initial and final syllable positions (e.g., /n/ and /l/).

Tables V and VI show the confusion matrices for the 34 560 initial and final consonants presented in the main experiment. The patterns of confusion resemble those reported in previous studies (Miller and Nicely, 1955; Wang and Bilger, 1973; Phatak and Allen, 2007). Across all consonants, FA rates in excess of 2% were produced by an average of 4.6 alternatives and FAs in excess of 1% were produced by an average of 7.2 alternatives. The remaining 11.8 response alternatives (particularly those differing in multiple phonetic

TABLE VI. Confusion matrices for final consonants averaged across subjects, SNRs, and voices. Each consonant was delivered on 1728 trials.

	b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k
b	1116	94	91	1	11	6	17	29	184	56	1	4	2	2	0	19	19	60	10	6
d	103	1114	135	9	12	10	38	12	97	87	6	36	0	0	2	29	7	16	9	6
g	122	91	1140	5	7	9	9	15	128	54	2	6	0	1	0	18	14	33	17	57
r	27	49	112	925	128	25	44	19	74	14	38	65	19	24	12	30	17	32	37	37
l	51	22	84	123	1026	33	48	42	141	24	7	23	2	6	2	16	16	20	24	18
ŋ	7	7	14	4	42	877	390	281	69	22	0	1	0	1	0	7	3	2	1	0
n	11	15	11	7	14	90	1327	159	54	26	0	3	1	0	0	4	4	2	0	0
m	17	7	6	2	28	78	182	1328	56	15	3	0	0	0	0	2	2	1	0	1
v	64	9	17	3	7	3	7	7	1468	131	3	0	0	0	0	3	4	1	1	0
ð	10	9	3	1	3	1	2	3	714	940	23	1	0	0	0	13	4	0	1	0
z	25	55	62	16	11	10	24	21	70	32	948	144	40	34	61	32	16	21	72	34
ʤ	21	68	66	0	11	13	11	14	38	13	15	1269	75	29	4	14	12	12	27	16
ʧ	9	3	21	3	2	6	6	9	8	6	10	285	1066	146	9	17	9	10	65	38
ʃ	6	6	29	2	4	14	15	7	17	8	13	221	214	1008	40	19	5	18	39	43
s	21	20	57	7	9	11	32	21	42	22	177	82	59	106	746	76	44	39	111	46
θ	1	2	2	0	0	0	1	0	26	19	3	0	0	3	9	1314	337	1	7	3
f	6	2	6	0	1	1	4	2	67	16	3	1	2	0	9	394	1178	9	10	17
p	112	12	21	3	6	1	8	14	18	7	1	2	11	5	0	32	49	1095	114	217
t	6	40	32	5	6	6	15	12	17	8	8	37	37	12	10	46	29	99	1171	132
k	16	12	59	2	8	5	12	11	10	4	4	9	15	8	4	60	48	149	90	1202

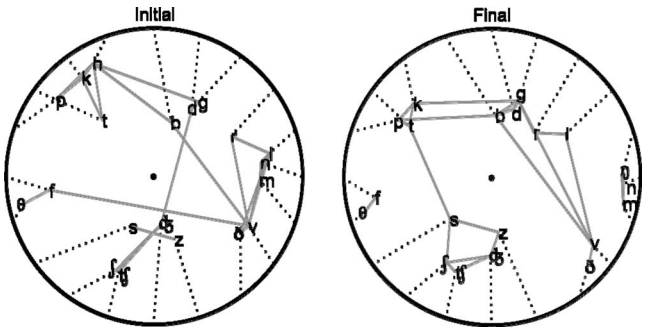


FIG. 2. Confusion clusters for initial and final consonants. Left: Consonant confusion clusters for initial consonants visualized using barycentric clustering. Initial consonants were first equally spaced around the confusion circle reflecting phonetic classification of voicing, manner, and place so that starting locations (clockwise from top center) were /b/, /d/, /g/, /r/, /l/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, /k/, and /h/, with /h/ adjacent to /b/. Then, each consonant was displaced from its initial position (dashed line) using iterated weighted averaging of its position with that of every other consonant based on the frequency of confusions between each consonant pair. Consonants that were frequently confused (e.g., /v/ and /ð/) cluster together at intermediate locations that also reflect confusions of one or both consonants in the pair with all other consonants. Gray lines connect the 12.5% of consonant pairs that showed greatest confusions. Right: Consonant confusion clusters for final consonants with starting position (top left-center) /b/, /d/, /g/, /r/, /l/, /ŋ/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, and /k/, with /k/ adjacent to /b/.

features from the target consonant) produced very low FA rates. Both FAs (in the columns) and misses (along the rows) were concentrated near the correct-response diagonal, indicating a preponderance of place errors. Other common confusions reflected manner errors conserving place and voice (e.g., /b/-/v/ and /b/-/m/ confusions) and voicing errors conserving place and manner (e.g., /b/-/p/ and /ʧ/-/ʤ/ confusions).

The confusion matrices presented in Tables V and VI reveal that consonants fell into confusable clusters of varying sizes. For some consonants, confusions were largely confined to consonants that shared manner (e.g., nasals in final consonant position). In contrast, other consonants showed confusions with larger consonant sets (e.g., initial /v/ with /b/, /ð/, /m/, /l/, /r/, and /f/). Figure 2 shows the clustering of consonant confusions, visualized using the barycentric algorithm developed by Cohen (2009) to characterize interrelated clusters of internet users. Consonants were initially placed in equidistant positions around a unit circle based on voicing, manner, and place features using the *a priori* consonant ordering shown in Tables V and VI. Unvoiced and voiced consonants were initially positioned on left and right sides of the circle, connected at the bottom across the voiced and unvoiced affricates and connected at the top across voiced and unvoiced plosives. Then, the location of each consonant was computed as an average of its current position, weighted by its hit rate, and the position of every other consonant weighted by number of false responses to that consonant. Two iterations were used to generate the cluster plots shown in Fig. 2. As a result of these iterations, each consonant was displaced from its initial position (dotted lines in Fig. 2) toward the locations of consonants with which it was confused.

Consonant pairs producing the most frequent confusions (top 12.5%, corresponding to total bi-directional confusion rates in excess of approximately 6% of hits) are shown connected by solid gray lines. For example, in final syllable position (Fig. 2, right) nasals were frequently confused with each other but rarely confused with any other consonants, so they moved to an intermediate position near /n/ on the circle edge. In contrast, in the initial syllable position (Fig. 2, left), the voiced sibilant /z/ was frequently confused with the unvoiced sibilant /s/ but also showed moderate rates of confusion with /t/, /ð/, /r/, and /d/, resulting in its displacement toward the center of the circle (Fig. 2, left).

The magnitude of the total displacement of consonants from their initial positions reflects both the initial consonant locations and the observed consonant confusions. For example, if two confusable consonants were initially placed in opposite positions on the confusion circle, both would move substantial distances toward the circle center. In contrast, if two confusable consonants were initially placed in adjacent positions, only small displacements would occur between their locations near the circle boundary. Thus, the overall magnitude of consonant displacement reflects in part the degree to which the initial consonant positions accurately predicted the confusions observed. In the current experiment, initial consonant placements were determined based on place, manner, and voicing features. Feature analysis (see below) revealed that place confusions occurred more frequently than manner confusions, and that manner confusions occurred more frequently than voicing confusions. Therefore, voiced and unvoiced consonants were placed on opposite sides of the circle. Within each voicing group, consonants were then segregated by manner. Finally, within each manner and voicing grouping, consonants were ordered by place of articulation. Consonants whose manner also defined that they are voiced (nasals and liquids) were centered within the voiced-consonant group. At the borders separating different manners, adjacent consonants were positioned to share place and voicing. Similarly, at the borders separating different voicings, adjacent consonants were positioned to share place and manner.

Permutation testing was used to evaluate how accurately the *a priori* consonant positions reflected the observed pattern of consonant confusions. This was accomplished by comparing the total consonant displacement distance with that produced by 10^7 unique random consonant positioning schemes. The *a priori* placement resulted in less displacement than 99.9998% of the random position schemes tested, with no significant differences observed between the magnitude of displacement obtained with the *a priori* positions and the magnitude of displacement obtained with the optimal random placement. This suggests that the initial consonant positions accurately captured the actual confusion clustering evident in the data.

Figure 2 shows the patterns of consonant confusions observed for initial and final consonants. Among initial consonants, the unvoiced plosives /p/, /t/, and /k/ were frequently confused with each other and also with /h/, and thus formed a relatively tight cluster in the upper right of the confusion circle. A relatively high frequency of /p/-/h/ confusions resulted in both /p/ and /h/ being attracted toward a location

intermediate between their initial positions. Because /t/ was occasionally confused with the sibilants and affricates located in the lower right quadrant of the circle, both /t/ and the sibilant-affricate cluster were displaced toward the circle center. These sibilant-affricate confusions were less common for /p/, /k/, and /h/, so these consonants remained closer to the circle circumference than /t/ in the unvoiced-plosive cluster.

Figure 2 shows clear confusion clustering of initial unvoiced plosives, voiced plosives, unvoiced non-sibilant fricatives, sibilants and affricates, voiced fricatives, and liquids and nasals. A few initial consonants (e.g., /v/, /b/, /h/, and /ʤ/) were frequently confused with consonants in more distant clusters. For example, voicing confusions were relatively frequent for /ʤ/-/ʧ/ pairs, and both of these consonants were confused with /ʃ/, /s/, and /z/. Similarly, voicing confusions occurred between /f/, /v/, /b/, and /h/, and manner confusions occurred between /b/ and /v/. In addition, /v/ was frequently confused with both nasals and liquids and was tightly clustered with /ð/, reflecting frequent confusions between this consonant pair. In general, confusion clusters were similar for initial and final consonants with a few exceptions. In the final consonant position, unvoiced plosives formed a tighter cluster in the absence of /h/, and nasals were particularly well discriminated from all other consonants including both voiced non-sibilant fricatives and liquids.

D. Analyzing phonetic features

Figure 3 shows the incidence of single-feature errors (place, manner, and voicing) and combined-feature errors for initial consonants (top) and final consonants (bottom) at the three different SNRs. Table VII shows the average number of consonants in each of the different feature-error categories.

ANOVA for repeated measures was first used to analyze error rates for different phonetic features with position, SNR, and feature (place-only [P], manner-only [M], voicing-only [V], P+M, P+V, M+V, and P+M+V) as factors. This analysis revealed no significant effect of position in the syllable, but revealed the expected main effect of SNR [$F(2,30)=1524.02$, $p<0.0001$]. There was also a large main effect of feature [$F(6,90)=467.28$, $p<0.0001$]: P errors were most frequent (10.2% of trials), followed by P+M (8.6%), M (6.6%), and V (3.3%) errors. The relatively high incidence of P+M errors has been reported previously (Dubno and Levitt, 1981) and reflected in part the fact that there were nearly twice as many consonants in the P+M category as the M category, and more than four times as many consonants in the P+M category as in the P category (Table VII). The remaining combined-feature errors generally occurred on less than 3% of trials except at the most difficult SNR. In addition, there were significant position × feature [$F(6,90)=85.85$, $p<0.0001$] and SNR × feature [$F(12,180)=138.01$, $p<0.0001$] interactions that were explored in additional ANOVAs.

The next ANOVA compared the incidence of different single-feature errors (P, M, and V). This analysis showed a main effect of feature [$F(2,30)=427.76$, $p<0.0001$], primarily reflecting the increased overall incidence of P errors

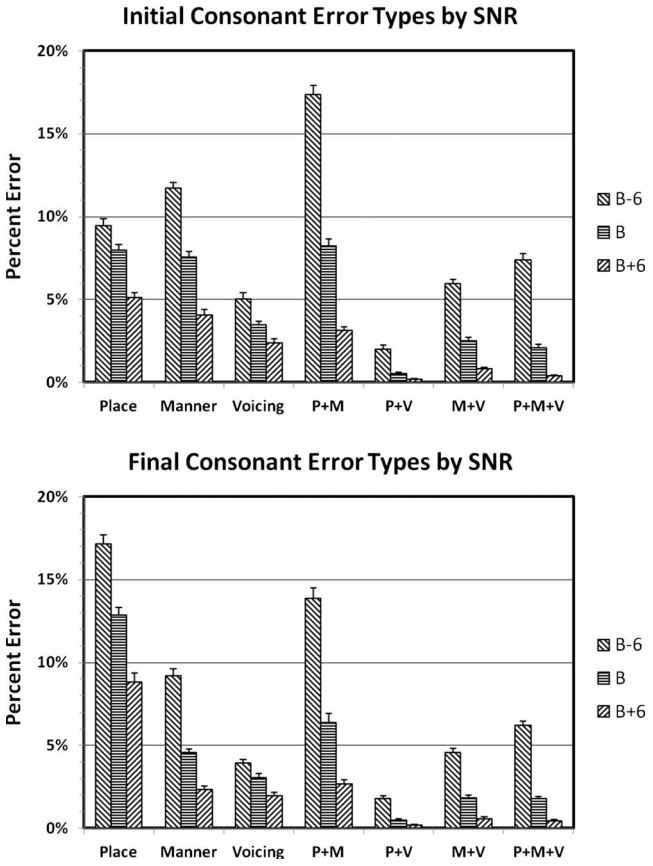


FIG. 3. Error rates for different phonetic features. Place, manner, voicing, P+M=place+manner, etc. Error bars show standard errors of the mean.

in comparison with M and V errors. There was also a main effect of position [$F(1,15)=22.18$, $p<0.0005$] due to fewer single-feature errors for initial than final consonants that was accompanied by a correspondingly greater incidence of combined-feature errors among initial consonants. Finally, there was a highly significant position × feature interaction [$F(2,30)=153.90$, $p<0.0001$], reflecting the fact that M and V errors were more common for initial than final consonants while P errors showed the opposite pattern.

In order to evaluate the changes in feature processing that accompanied increases in SNR, error rates for each feature and position were normalized relative to the total number of errors of each type. The SNR × feature interaction was significant [$F(12,180)=96.13$, $p<0.0001$], indicating that SNR differentially affected error rates for different features. An additional specific analysis examining only single-feature errors showed a significant effect of feature [$F(2,30)=53.02$, $p<0.000\,01$], reflecting the fact that M and V errors declined more rapidly than P errors with increases in SNR. Finally, additional *post-hoc* testing demonstrated that combined-feature errors declined more rapidly than P, M, or V errors ($p<0.001$ for all comparisons).

E. Vowel influences on consonant identifiability

The effects of vowel context on consonant processing were analyzed by comparing consonant identification performance as a function of the accompanying vowel for the 19 consonants that appeared in both initial and final syllable

TABLE VII. Average number of consonants sharing different features and combinations of features with each target consonant.

	Place	Manner	Voicing	P+M	P+V	M+V	PMV
Initial	1.30	2.50	0.70	5.30	1.10	2.60	5.50
Final	1.30	2.50	0.70	5.55	0.90	2.60	5.40

positions. The results were analyzed using repeated measures ANOVA with subjects, vowel, position, and consonant as factors. The SNRs of syllables containing different vowels had been adjusted in preliminary studies to reduce vowel-specific differences in overall mean syllable identifiability. As a result, the vowel main effect failed to reach significance [$F(2,30)=2.62$, $p<0.10$]. However, there was a highly significant vowel×position interaction [$F(2,30)=42.40$, $p<0.0001$]. This reflected the fact that initial consonants were identified more accurately than final consonants in syllables containing /ɑ/, whereas final consonants were identified more accurately than initial consonants in syllables containing /i/. Initial and final consonants were identified with comparable accuracy in syllables containing /u/.

There was also a highly significant vowel×consonant interaction [$F(36,540)=45.67$, $p<0.0001$], as shown in Table VIII. Separate ANOVAs analyzing vowel effects for each of the 19 consonants showed that d' measures for 16 of the 19 consonants (excepting /ʧ/, /d/, and /k/) were significantly influenced by the accompanying vowel. Vowel-related changes in d' ranged from 0.37 to 1.46 for different consonants and were generally highly significant [the average over all consonants was $F(2,30)=44.23$, $p<0.0001$]. For example, /m/ was identified much more accurately in syllables containing /ɑ/ than in syllables containing /i/ or /u/, producing vowel-related d' differences of 1.46. This would be equivalent to a 9.1 dB change in SNR estimated based on the P/S slope for /m/ ($0.16d'$/dB).

Figure 4 shows identification performance for initial and final consonants displayed on a vowel triangle with the distance of each consonant from each vowel apex inversely scaled by its d' value when presented with that vowel. Consonants with similar d' values for all vowels are located near the center of the triangle (e.g., /ʧ/). Figure 4 shows that initial nasals were preferentially identified with a /ɑ/, and better identified with /u/ than /i/, while final nasals were preferentially identified with a /ɑ/, but better identified with /i/ than /u/. Initial non-sibilant fricatives were poorly identified with /i/ and therefore clustered on the /ɑ/-/u/ boundary. In final syllable position, the fricatives /v/, /f/, and /θ/ showed less vowel specificity, while /ð/ was very poorly identified with both /a/ and /i/. Similarly, /t/ was poorly identified with /ɑ/ in both consonant positions, while /r/ was poorly identified with /u/ as an initial consonant and poorly identified with /ɑ/ as a final consonant.

Confusion matrices for initial and final consonants in syllables containing different vowels are included in Supplementary Tables 1–6 (see Supplementary material). Figure 5 shows the cluster confusion analyses as a function of vowel. An examination of the vowel effects on the confusion matrices clarifies the nature of the vowel-related changes in consonant confusions. For example, the confusion clusters of nasal consonants (/m/, /n/, and in final consonant position, /ŋ/) show that initial nasals are moderately confused with each other and also with the liquid /l/ when presented with /ɑ/. In contrast, when initial nasals were presented with /i/ or /u/ they were more often confused with each other (i.e., they are closer together in the confusion circle) and also confused with /l/, /r/, /v/, and /ð/. Similarly, final nasals presented with /ɑ/ are tightly clustered near the circle periphery, reflecting a low incidence of confusions with non-nasal consonants. They are also widely spaced from each other, reflecting enhanced within-nasal discrimination. Final nasals presented with /i/ remain distinct from other confusion clusters, but show increased within-nasal confusions. Final nasals presented with /u/ show a high frequency of within-nasal confusions as well as additional confusions with /l/ and /v/. Similarly, the poor discriminability of the final /ð/ when presented with /ɑ/ and /i/ can be seen to reflect its very poor discriminability from /v/. The discriminability of /v/ and /ð/ is improved when these consonants are presented with /u/.

Additional analyses were performed to examine the effects of vowel context on the processing of consonants with different manners, with vowel, position, and manner as factors. This analysis revealed a highly significant interaction between vowel and manner [$F(8,120)=125.44$, $p<0.0001$]. Sibilants, liquids, and plosives were identified most accurately in syllables containing /i/, non-sibilant fricatives were identified most accurately in syllables containing /u/, and nasals were identified most accurately in syllables containing /ɑ/.

The next analysis examined vowel context effects on

TABLE VIII. d' scores for initial and final consonants in syllables containing different vowels.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	/ɑ/	2.24	1.90	2.25	2.23	2.72		3.04	3.12	1.97	2.35	1.88	1.67	2.27	2.16	1.96	2.41	2.33	2.16	1.57	2.41	2.28
	/i/	2.20	2.31	2.13	2.51	2.00		1.45	1.62	1.60	1.52	2.42	1.86	2.08	2.41	2.21	2.01	1.85	2.94	2.28	2.38	2.38
	/u/	1.94	2.28	2.04	1.74	1.78		2.27	2.01	2.27	2.48	1.95	1.98	2.22	2.34	2.06	3.08	2.95	2.17	2.20	2.19	2.03
Final	/ɑ/	2.34	2.18	1.98	1.82	2.33	2.56	2.86	3.32	2.11	1.46	1.97	1.88	2.02	1.86	1.81	2.08	1.92	2.00	1.77	1.88	
	/i/	2.01	2.01	2.42	2.69	2.87	2.32	2.02	2.25	2.05	1.61	2.38	2.32	2.09	2.33	2.43	2.27	2.20	2.38	2.28	2.31	
	/u/	1.82	2.21	1.78	2.49	1.90	1.58	2.16	1.97	2.35	2.82	2.20	2.19	2.25	2.31	2.02	2.81	2.40	2.12	2.31	2.26	

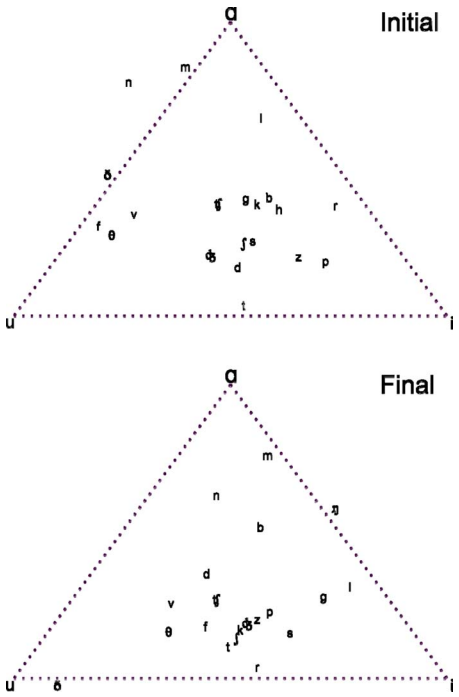


FIG. 4. (Color online) Consonant-vowel interactions shown on a triangular map of vowel space with the three vowels presented in the current experiment at the apices. The distance of each consonant from each vowel apex is inversely related to the d' measure for that vowel-consonant combination. Top=initial consonants; bottom=final consonants. The few consonants without significant vowel preferences (e.g., /ʧ/) are found near the center of the triangle. Consonants with strong vowel preferences lie closer to the preferred vowel (e.g., nasals are close to /ɑ/).

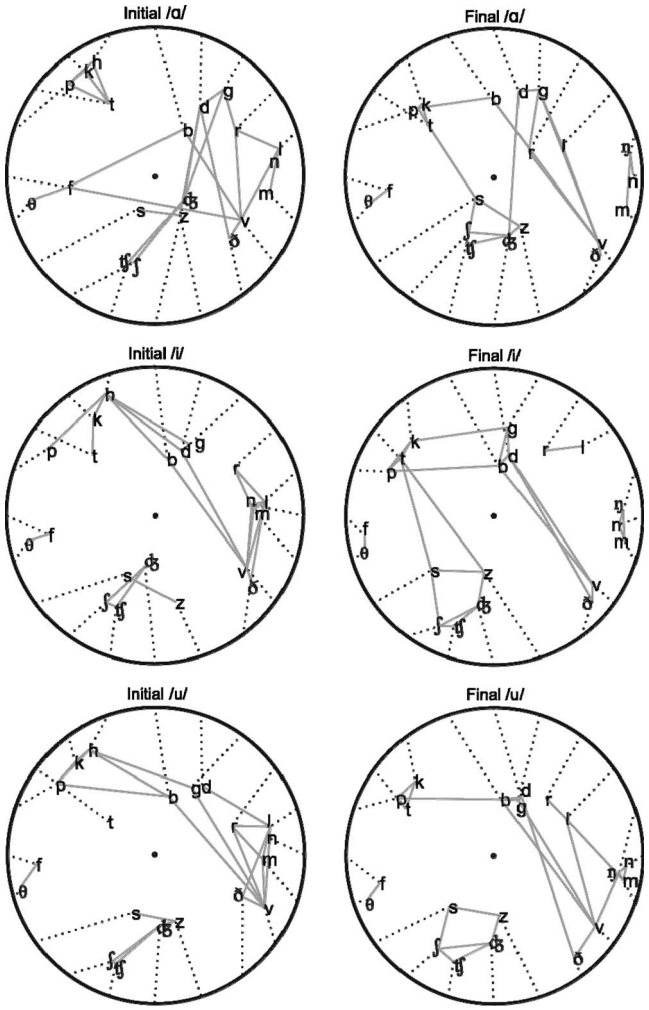


FIG. 5. Confusion clusters for initial (left) and final (right) consonants as a function of vowel. See Fig. 2 for further details.

place of articulation. There was a highly significant vowel $\times$place interaction $[F(4,60)=34.58,\ p<0.0001]$. Front consonants were perceived most accurately in syllables containing /ɑ/, middle consonants were perceived most accurately in syllables containing /u/, and back consonants were perceived most accurately in syllables containing /i/. There was also a significant interaction between vowel, syllable position, and place $[F(4,60)=6.02,\ p<0.002]$, indicating that vowel context effects on place differed at initial and final syllable positions.

Finally, there was also a significant vowel$\times$voicing interaction $[F(2,30)=57.97,\ p<0.0001]$. Unvoiced consonants were most accurately identified in syllables containing /u/, whereas voiced consonants were most accurately identified in syllables containing /ɑ/. There was no significant difference in vowel effects on the voicing of consonants at initial and final syllable positions $[F(2,30)=0.93, \text{NS}]$.

Vowel effects on feature processing were analyzed by evaluating the relative incidence of various types of feature errors in the vowel-specific confusion matrices presented in Supplementary Tables 1–6 (see supplementary material). An omnibus ANOVA indicated a highly significant interaction between vowel and error-type (P, M, V, P+M, P+V, M+V, and PMV) $[F(12,180)=33.70,\ p<0.0001]$ that was explored in additional ANOVAs. There were no significant main effects of vowel on the incidence of V errors $[F(2,30)=0.12]$ and only a trend toward a reduction in M errors with the vowel /i/ $[F(2,30)=3.36,\ p<0.06]$. However, there was a highly significant vowel main effect on the incidence of P errors $[F(2,30)=78.52,\ p<0.0001]$. P errors were increased in syllables containing /i/ relative to those containing /ɑ/ and /u/. P errors also showed a significant vowel$\times$position interaction $[F(2,30)=21.03,\ p<0.0001]$. Inter-vowel differences increased in the final syllable position due in large part to the overall increase in P errors in the final syllable position described above. An analysis of combined-feature errors also revealed a significant vowel main effect $[F(2,30)=53.81,\ p<0.0001]$ on the relative incidence of P+M errors. P+M errors were reduced in syllables containing /i/ relative to those containing /ɑ/ and /u/, without a significant vowel$\times$position interaction.

F. Interactions in identifying initial and final consonants

We evaluated the effects of consonant context using repeated measures ANOVAs with subject, vowel, consonant position, and matching of initial and final consonant features as factors. Separate analyses were performed for manner, place, and voicing features, each evaluated at the $p<0.01$ level of significance. There were no significant effects of matching on manner $[F(1,15)=1.58]$ or voicing $[F(1,15)=3.99,\ p<0.07]$. However, there was a highly significant

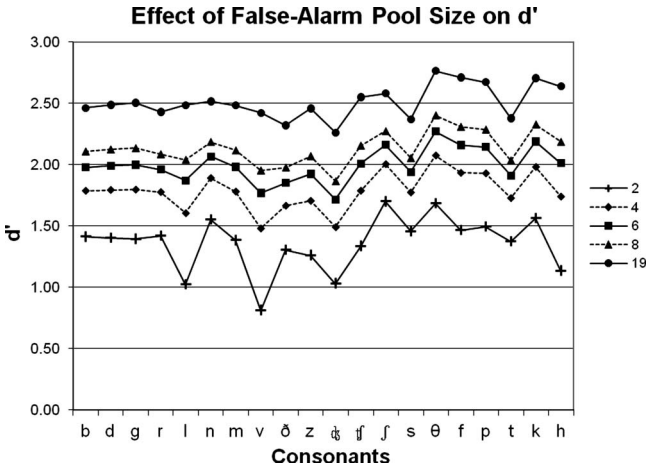


FIG. 6. d' values for initial consonants calculated with different false-alarm-pool sizes.

effect of matching on place $[F(1,15)=41.70,\ p<0.0001]$: subjects showed increased accuracy when initial and final consonants differed in place of articulation. This effect did not differ as a function of consonant position.

IV. DISCUSSION

A. Quantifying consonant identification

Both hit rate and d' were examined as metrics of consonant identification performance. Hit rates were confounded by large differences in response criteria associated with nearly ten-fold differences in the FA rates of different consonant responses. Miller and Nicely (1955) found similar results: up to ten-fold differences in the FA rates for different consonants at certain SNRs. Therefore, d' appeared to be a better metric for evaluating consonant identifiability because of its correction for differences in response criteria.

Unfortunately, there is no standard method of calculating d' under conditions where many potentially confusable signals are presented in noise. In the standard signal detection theory paradigm, the only trials presented are signal-plus-noise (associated with hits) or noise-alone (associated with FAs). However, in the current study, FAs assigned to a particular consonant response were usually generated by a number of different consonants with which it shared phonological and acoustic features. For example, /v/ FAs occurred on 29% of /ð/ trials, but on less than 0.2% of ʃ trials. These results suggest that some, but not all, consonants could be sensibly considered plausible FA-producing distractors. In our analysis we included only the plausible FA-producing consonants in the FA pool used to determine N, the divisor for the FA rate calculation. As noted in Sec. II, the size of the FA pool was estimated to be 6.5 consonants based on preliminary studies. This FA-pool size corresponded to consonants producing FA rates in excess of approximately 1.3% and was slightly greater than a FA-pool size (5.3) based on consonants that shared two phonetic features (e.g., manner and voicing) with each response.

Figure 6 shows d' values for initial consonants calculated using different FA-pool sizes ranging from 2 to 19 consonants. Increasing the FA-pool size decreases in FA rate and correspondingly increases the d'. The increase in d' values is greatest for consonants with high FA rates, e.g., /v/, /l/, and /h/ in Fig. 6. Although the d' curves differ somewhat at extreme FA-pool sizes, they have similar shapes over intermediate FA-pool sizes, indicating that the current results are largely insensitive to changes in the FA-pool size over realistic intermediate ranges.

B. Consonant identifiability in noise

The current experiment assessed listeners' abilities to identify 20 initial and 20 final consonants using spoken responses that included the majority of potential consonant confusions. To assure that confusion patterns were representative of average articulation patterns, syllables were randomly sampled from a set of 9600 tokens with equal sampling of each talker, vowel, and consonant at each SNR. Because the results of the current study are based on the random sampling of a very large CVC token set, they provide insights into the identification of American English consonants in typically articulated CVCs.

It is well known that some consonants are much easier to identify in noise than others (Wang and Bilger, 1973; Dubno and Levitt, 1981; Dubno *et al.*, 1982; Helfer and Huntley, 1991; Gelfand *et al.*, 1992; Benkí, 2003; Cutler *et al.*, 2004; Phatak and Allen, 2007; Phatak *et al.*, 2008). Previous studies suggest that SNR levels must be adjusted by 18 dB (Miller and Nicely, 1955) to 24 dB (Phatak and Allen, 2007) to equate hit rates across all consonants in 16-consonant sets. We found that even larger differences in SNR were needed to equate consonant identifiability in a 20-consonant set of initial (28.8 dB) and final (47.0 dB) consonants. The increased range of SNRs needed to equate consonant identifiability likely reflects in part the increased number of possible consonant confusions. The addition of the consonants /l/, /r/, /ʧ/, /h/, and /ŋ/ in the 20-consonant sets increased potential confusions for many consonants (particularly /ð/), reducing their discriminability and correspondingly increasing their required baseline SNR levels.

The pattern of results shown in Table III suggests that a subset of consonants (sibilants, affricates, liquids, and some plosives) provides the majority of consonant information in difficult listening situations (e.g., absolute SNRs less than 6 dB). Consonant information becomes available from less easily identified consonants only at higher SNRs. The results also revealed surprisingly large differences in identifiability of different consonants within the same phonetic class. For example, among initial plosives, baseline SNRs for /p/ and /t/ differed by 12.4 dB, while baseline SNRs for /b/ and /d/ differed by 7.1 dB. Phatak and Allen (2007) observed similar differences in the relative identifiability of these consonant pairs. These differences in plosive identifiability may reflect the greater high-frequency spectral energy in /t/ and /d/ plosive bursts with respect to /p/ and /b/ (Halle *et al.*, 1957), resulting in less effective masking by low-frequency speech-spectrum noise of the sort used in the current study.

Although the average psychometric functions obtained in this experiment had mean slopes that were similar to those reported in previous studies using nonsense CVCs (Boo-

throyd and Nittrouer, 1988), there were large differences in the P/S slopes of different consonants. Not surprisingly, harder-to-identify consonants (i.e., those that required higher baseline SNRs) had shallower slopes. This was due in part to the fact that harder-to-identify consonants (e.g., non-sibilant fricatives) remain difficult to identify even at high SNRs (Wiley *et al.*, 1979). In contrast, easier-to-identify consonants (e.g., sibilants) showed steeper psychometric functions. The correlation between baseline SNRs and P/S slopes suggests that consonant information from easy-to-identify consonants continues to provide the majority of consonant information in conversational speech at moderate noise levels.

C. The identification of leading and trailing consonants

Overall, we found that initial consonants were identified at consistently lower SNRs than final consonants, replicating the results of previous studies using smaller token sets (Wang and Bilger, 1973; Dubno and Levitt, 1981; Redford and Diehl, 1999; Benkí, 2003). These differences likely reflected slight increases in the amplitude and duration of initial consonants with respect to final consonants (Redford and Diehl, 1999). However, the magnitude of the initial consonant advantage differed for different consonant manners. Nasals showed particularly large initial consonant advantages as in previous reports (Dubno and Levitt, 1981; Repp and Svastikula, 1988). In contrast, liquids showed a final consonant advantage that has likewise been previously noted (Cutler *et al.*, 2008). Although d' had been equated for initial and final consonants, we found a significant position-related difference in response criteria: initial consonants were identified with stricter criteria than final consonants. In part, this may have reflected differences in the 20th consonant used in initial and final consonant sets: /h/ was a more potent FA-generator than /ŋ/.

Cluster analysis was used to visualize the confusions among consonants in initial and final syllable positions. This analysis revealed that the particular *a priori* categorization of consonants based on voice, manner, and place features provided a valid starting point for consonant-confusion analysis. Although consonants generally segregated into confusable clusters sharing manner and voicing, cluster analysis revealed considerable overlap between nasals and liquids and also between sibilants and affricates. Distant clusters were connected through "hub" consonants (e.g., /h/, /v/, /b/, and /ʤ/) that were frequently confused with consonants with distant clusters characterized by different manner and/or voicing features. Cluster analysis showed that confusion clusters were similar in initial and final consonant positions, with some clusters better segregated in initial syllable position and other clusters better segregated in final syllable position.

D. Analyzing phonetic features

Phonetic features differed in intelligibility. Among single-feature errors, P confusions were most common, followed by M and V confusions. Moreover, P confusions persisted to a greater degree at higher SNRs than M and particularly V confusions as in previous reports (Grant and Walden, 1996). We also found a high incidence of P+M errors, particularly at low SNRs, consistent with previous studies (Dubno and Levitt, 1981). P+M errors and other combined-feature errors declined more rapidly than did single-feature errors with increasing SNR.

The relative salience of different phonetic features also differed significantly as a function of syllable position. Combined feature errors were more common in initial than final consonant position (possibly because of the inclusion of /h/), while single-feature errors showed the opposite pattern. Moreover, among single-feature errors, V and M confusions were more common among initial consonants while P confusions were more common among final consonants. Differences in the articulation of initial and final consonants may have contributed to these results (Redford and Diehl, 1999). In addition, there are differences in the perceptual operations associated with the identification of initial and final consonants. Subjects begin extracting information about the initial consonant prior to receiving vowel information and must identify the consonant and vowel concurrently. The parallel analysis of the initial consonant and vowel may increase multi-feature as well as M and V errors that depend on the accurate analyses of the vowel nucleus. In contrast, the increase in place errors in final consonant position may have reflected the relative reduction in M and V errors in the final position subsequent to vowel identification. Alternatively, uncertainty in the onset time of the initial consonant and the resulting failure to accurately detect its onset might have given rise to increased multi-feature errors. The temporal uncertainty of final consonant delivery would have been reduced because of the additional timing cues provided by the vowel and the start of the vowel-consonant transition.

E. Vowel influences on consonant identifiability

Overall differences in consonant identifiability in syllables containing different vowels had been eliminated by small vowel-specific adjustments in syllable intensity. However, different vowel effects were observed in initial and final syllable positions. Initial consonants were more accurately identified in syllables containing /ɑ/ while final consonants were more accurately identified in syllables containing /i/, as observed in previous reports (Halle *et al.*, 1957; Wang and Bilger, 1973; Redford and Diehl, 1999).

Further analysis revealed that vowel space was not independent of consonant space: highly significant vowel × consonant interactions for 85% of consonants. These interactions were generally similar for consonants sharing similar manner. For example, nasals were more accurately identified with /ɑ/, as observed previously (Dubno and Levitt, 1981; Repp and Svastikula, 1988), while sibilants, plosives, and liquids were most accurately identified in syllables containing /i/, and non-sibilant fricatives were most accurately perceived in syllables containing /u/. These effects may have reflected more salient formant transitions in preferred consonant-vowel pairs. Vowels also interacted with place of articulation. Front consonants were perceived most accurately in syllables containing the back vowel /ɑ/, middle consonants were perceived most accurately in syllables contain-

ing /u/, and back consonants were perceived most accurately in syllables containing the front vowel /i/. These effects may have reflected the increased salience contrasts associated with greater tongue movements and presumably longer duration formant changes.

Confusion-cluster analysis was used to visualize vowel-consonant interactions. This analysis revealed variations in the clarity of different phonetic cues as a function of the vowel-nuclei of syllables. Vowel-related improvements in consonant identification were usually accompanied by a reduction in place confusions within clusters as well as reductions in confusions with distant clusters. We also found systematic differences in the overall frequency of P confusions as a function of vowel: P confusions increased with /i/ in comparison with /ɑ/ or /u/, whereas P+M errors showed a corresponding decrease. One possible explanation is that /i/ facilitated manner processing and thus converted potential P+M errors into P errors.

F. Interactions in identifying initial and final consonants

We found evidence of significant interactions between the processing of initial and final consonants in CVCs: both initial and final consonants were identified more accurately when they differed in place of articulation. No such interactions were found for voicing or manner. One possible explanation is because the analysis of place is more difficult than the analysis of manner or voicing, particularly at the final consonant position, it requires more time for analysis. This would increase the possibility of consonant-consonant interactions in place processing over the relatively long inter-consonant intervals of CVCs (Christiansen and Greenberg, 2008). The reduced accuracy found when initial and final consonants share place may thus reflect adaptation of the place feature at relatively short intervals.

V. CONCLUSIONS

Consonant identification in CVCs presented in noise is a highly complex process that is subject to many contextual influences. Different consonants are identifiable over widely different SNR ranges, suggesting that a subset of consonants contributes disproportionately to speech understanding during difficult listening situations. Cluster analysis of confusions revealed that consonants are segregated into perceptually distinct clusters, exhibiting high intracluster confusion rates and generally low rates of intercluster confusions. Perceptual clusters typically consist of consonants with the same manner and voicing, with the exception of nasal-liquid and sibilant-affricate clusters. Highly significant differences were observed in the identification of initial and final consonants, and consonant identification was strongly modulated by the accompanying vowel. In addition, consonant-consonant interactions were observed in the processing of initial and final consonant place features. These results suggest that consonant processing within CVCs is a complex, non-linear process that is subject to contextual modulation by consonant position, intervening vowels, and even other consonants.

[1]Pure tone thresholds (over 500, 1000, and 2000 Hz) averaged 6.71 dB HL (±1.93) in the left ear and 4.80 dB HL (±1.52) in the right, with average 8000 Hz thresholds of 10.94 dB HL (±2.95) and 6.88 dB HL (±2.58) in the left and right ears, respectively.

[2]Experimenter response transcription was used in preference to subject transcription to maintain the naturalness of the listening task, minimize procedural learning effects, and avoid scoring biases that might be introduced by listeners untrained in the use of the phonetic alphabet.

Benkí, J. R. (**2003**). "Analysis of English nonsense syllable recognition in noise," Phonetica **60**, 129–157.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Christiansen, T. U., and Greenberg, S. (**2008**). "Cross-spectral synergy and consonant identification," J. Acoust. Soc. Am. **123**, 3850.

Cohen, J. (**2009**). "Graph twiddling in a MapReduce world," Comput. Sci. Eng. **11**, 29–41.

Cutler, A., Garcia Lecumberri, M. L., and Cooke, M. (**2008**). "Consonant identification in noise by native and non-native listeners: Effects of local context," J. Acoust. Soc. Am. **124**, 1264–1268.

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.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired listeners using a nonsense-syllable test. II. Syllable recognition and consonant confusion patterns," J. Speech Hear. Res. **25**, 141–148.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.

Gelfand, S. A., Schwander, T., Levitt, H., Weiss, M., and Silman, S. (**1992**). "Speech recognition performance on a modified nonsense syllable test," J. Rehabil. Res. Dev. **29**, 53–60.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Robert E. Krieger, Huntington, NY).

Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.

Halle, M., Hughes, G. W., and Radley, J. P. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

Helfer, K. S., and Huntley, R. A. (**1991**). "Aging and consonant errors in reverberation and noise," J. Acoust. Soc. Am. **90**, 1786–1796.

Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. **33**, 149–155.

Levitt, H., and Resnick, S. B. (**1978**). "Speech reception by the hearing-impaired: Methods of testing and the development of new tests," Scand. Audiol. Suppl. **6**, 107–130.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

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

Müsch, H., and Buus, S. (**2001a**). "Using statistical decision theory to predict speech intelligibility. I. Model structure," J. Acoust. Soc. Am. **109**, 2896–2909.

Müsch, H., and Buus, S. (**2001b**). "Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920.

Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.

Phatak, S. A., Lovitt, A., and Allen, J. B. (**2008**). "Consonant confusions in white noise," J. Acoust. Soc. Am. **124**, 1220–1233.

Redford, M. A., and Diehl, R. L. (**1999**). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," J. Acoust. Soc. Am. **106**, 1555–1565.

Repp, B. H., and Svastikula, K. (**1988**). "Perception of the [m]-[n] distinction in VC syllables," J. Acoust. Soc. Am. **83**, 237–247.

See supplementary material at http://dx.doi.org/10.1121/1.3293005 E-JASMAN-127-034003 for confusion matrices for individual vowels.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Wiley, T. L., Strennen, M. L., and Kent, R. D. (**1979**). "Consonant discrimination as a function of presentation level," Audiology **18**, 212–224.
Woods, D. L., and Elmasian, R. (**1983**). "The habituation of human event-related brain potentials elicited by speech sounds and tones," Soc. Neurosci. Abstr. **9**, 365.
Woods, D. L., Yund, E. W., and Herron, T. J. (**2010**). "Measuring consonant identification in nonsense syllables, words and sentences," J. Rehab. Res. Dev. **47** (In Press).

Learning to perceptually organize speech signals in native fashion[a)]

Susan Nittrouer[b)] and Joanna H. Lowenstein
Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Road, Suite 4000, Columbus, Ohio 43212

(Received 12 June 2009; revised 30 October 2009; accepted 31 December 2009)

The ability to recognize speech involves sensory, perceptual, and cognitive processes. For much of the history of speech perception research, investigators have focused on the first and third of these, asking how much and what kinds of sensory information are used by normal and impaired listeners, as well as how effective amounts of that information are altered by "top-down" cognitive processes. This experiment focused on perceptual processes, asking what accounts for how the sensory information in the speech signal gets organized. Two types of speech signals processed to remove properties that could be considered traditional acoustic cues (amplitude envelopes and sine wave replicas) were presented to 100 listeners in five groups: native English-speaking (L1) adults, 7-, 5-, and 3-year-olds, and native Mandarin-speaking adults who were excellent second-language (L2) users of English. The L2 adults performed more poorly than L1 adults with both kinds of signals. Children performed more poorly than L1 adults but showed disproportionately better performance for the sine waves than for the amplitude envelopes compared to both groups of adults. Sentence context had similar effects across groups, so variability in recognition was attributed to differences in perceptual organization of the sensory information, presumed to arise from native language experience. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3298435]

PACS number(s): 43.71.Ft, 43.71.Hw, 43.71.An [MAH] Pages: 1624–1635

I. INTRODUCTION

The ability to recognize speech involves sensory, perceptual, and cognitive processes. Regarding the first of these, listeners must have some amount of access to the acoustic signal generated by the moving vocal tract of the speaker. Sensory impairments such as hearing loss can inhibit speech recognition if the impairment sufficiently precludes access to that sensory information. Research involving listeners with hearing loss has commonly focused on this aspect of perception, with questions explicitly addressing how much of the signal and what properties are needed for speech perception (e.g., Dorman *et al.*, 1985; Erber, 1971; Leek *et al.*, 1987; Revoile *et al.*, 1985; Stelmachowicz *et al.*, 1990). But while having access to sensory information in the speech signal is an obvious prerequisite to recognition, it is not sufficient to explain speech perception. This was demonstrated by Surprenant and Watson (2001), who tested the speech recognition of college students and also examined their discrimination for a number of related non-speech stimuli. Performance on the two kinds of tasks were not strongly correlated, leading the authors to conclude that "…factors higher in the sequence of processing than the auditory periphery account for significant variance in speech recognition—by both normal hearing and by hearing-impaired listeners." (p. 2094). The experiment reported here was designed to test the hypothesis that part of that variance is explained by how listeners perceptually organize the sensory information in the speech signal.

The idea that perceivers need to organize the sensory information reaching them is well accepted in the study of visual perception where displays such as the classic Ruben's vase illustrate the phenomenon: In that demonstration, viewers recognize either a vase or two symmetrical profiles depending on how they organize the sensory information that reaches their retinas. Scientists studying general auditory perception similarly examine the principles that underlie the recognition of the object responsible for generating the sensory information (Binder *et al.*, 2004; Bregman, 1990; Dau *et al.*, 2009; Griffiths and Warren, 2004). When it comes to speech perception, however, considerably less attention has been paid to how listeners perceptually organize the components of the signal in order to recover the sound-generating object, perhaps due to long-standing controversy over whether listeners actually ever do recover that object, which is the moving vocal tract (e.g., Liberman *et al.*, 1962; Lotto *et al.*, 2009). Be that as it may, the work that has been done on perceptual organization for speech demonstrates that listeners must integrate disparate spectral and temporal signal components appropriately in order to recover cohesive and linguistically relevant percepts (e.g., Best *et al.*, 1989; Remez *et al.*, 1994).

The work reported here follows years of investigation examining how the weighting of explicit acoustic cues to phonetic identity changes as children acquire experience with a native language. Acoustic cues are defined as spectrally discrete and temporally brief pieces of the speech sig-

[a)]Portions of this work were presented at the 154th Meeting of the Acoustical Society of America, New Orleans, LA, November 2007.
[b)]Author to whom correspondence should be addressed. Electronic mail:nittrouer.1@osu.edu

 0001-4966/2010/127(3)/1624/12/$25.00

TABLE I. Mean weighting coefficients for formant transitions and fricative-noise spectra for listeners in each age group. Group numbers are in italics, standard deviations are in parentheses (from Nittrouer and Lowenstein, 2009).

	3-year-olds (*23*)	5-year-olds (*22*)	7-year-olds (*20*)	Adults (*20*)
Formant transition	0.50 (0.15)	0.48 (0.19)	0.32 (0.15)	0.30 (0.17)
Fricative noise	0.66 (0.13)	0.71 (0.14)	0.81 (0.11)	0.82 (0.11)

nal that affect listeners' phonetic labeling when modified (Repp, 1982). Several cues can underlie a single phonetic distinction, and the way those cues are weighted varies depending on native language experience (e.g., Beddor and Strange, 1982; Crowther and Mann, 1994; Flege *et al.*, 1996; Gottfried and Beddor, 1988; Zampini *et al.*, 2000). Accordingly, children must discover the optimal weighting strategies in their native language, and their early weighting strategies differ from those of adults. In particular, children weight dynamic, spectral cues (i.e., formant transitions) more than adults, and weight stable spectral cues (e.g., noise spectra) less (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b; Nittrouer and Studdert-Kennedy, 1987; Watson, 1997). Some evidence of this developmental shift comes from a study in which adults and children between 3 and 7 years of age were asked to label syllable-initial fricatives based on the spectral shape of fricative noises and on formant transitions at voicing onset (Nittrouer and Lowenstein, 2009). Computed weighting coefficients are shown in Table I. They indicate that the weighting of formant transitions diminished with increasing age, while the weighting of fricative-noise spectra increased.

Of course, formant transitions are just brief sections of more broadly modulating resonant frequencies created by vocal-tract cavities, which are continually changing in shape and size. Results across studies demonstrating children's reliance on formant transitions for phonetic labeling led to the broader hypothesis that children would show a perceptual advantage in organizing this kind of signal structure over other kinds of structure. To test that hypothesis, Nittrouer *et al.* (2009) presented two kinds of processed signals to three groups of listeners. Sine wave replicas of speech (Remez *et al.*, 1981) were used to provide a signal condition that preserved dynamic spectral structure, just the sort of structure for which children are expected to show an advantage. Vocoded stimuli (Shannon *et al.*, 1995) were also used as a comparison condition. These stimuli preserve amplitude structure across the utterance in a preselected number of frequency bands. The sentences in that study were all four words long with English syntax, but no semantic constraints. Two groups of adult listeners participated: native speakers of American English and native speakers of another language who were excellent second-language (L2) speakers of American English. Although native speakers of any language could presumably have been used, Mandarin Chinese speakers were selected because Fu *et al.* (1998) showed that they recognize vocoded Chinese sentences as well as native English speakers recognize vocoded English sentences. This group of adults was included both to test the idea that the perceptual organization of sensory information is language specific, and to serve as controls for the third group of listeners: 7-year-old native speakers of American English. If native English-speaking 7-year-olds had poorer recognition than L1 adults for either kind of processed stimuli, the inclusion of L2 adults would mitigate against the explanation that age-related differences in processing at the auditory periphery accounted for the outcome, assuming that L2 adults also performed more poorly than L1 adults.

Results showed that the L2 adults performed more poorly with both kinds of stimuli than the L1 adults, suggesting that how listeners perceptually organize acoustic structure in the speech signal is at least somewhat dependent on native language experience. That interpretation was motivated by outcomes of other studies. For example, Boysson-Bardies *et al.* (1986) showed that the long-term spectral shape of speech signals varies across languages, with some having more prominent spectral peaks and others having flatter spectra. The fact that listeners are sensitive to the spectral shape of their native language was demonstrated by Van Engen and Bradlow (2007), who reported that speech-shaped noise degraded speech recognition more when the shape of the noise was derived from speakers of the language in which testing was occurring, rather than from a different language. Apparently the way in which speech is produced creates structure in signals at global levels (i.e., longer than the phonetic segment and spectrally broad), and listeners are sensitive to how signals are structured at those levels in their native language.

The 7-year-old children in Nittrouer *et al.* (2009) were poor at recognizing sentences that preserved only amplitude structure, scoring similarly to the L2 adults. At the same time, 7-year-olds performed similarly to the L1 adults with the sine wave replicas. In fact, the 7-year-olds were the only group of listeners to show better performance for the sine wave stimuli than for the four-channel amplitude envelope (AE) stimuli (25.65% vs 13.68% correct, respectively); L1 and L2 adults had statistically identical recognition scores for the four-channel amplitude envelope and the sine wave stimuli: 28.17% correct for L1 adults across the two conditions and 9.99% correct for L2 adults. Of course, 7-year-olds' performance could have been rendered equivalent across conditions by increasing the number of channels in the amplitude envelope stimuli, but the critical outcome was precisely that they performed differently on two sets of data that evoked similar responses from adults. However, a caveat in those results was that different listeners participated in testing with the two kinds of signals, so within-group comparisons could not be made. That problem was corrected in the current study.

The perceptual advantage demonstrated by 7-year-olds in that experiment for sine wave stimuli was viewed as complimentary to earlier findings showing that children weight formant transitions more than adults in phonetic decisions. Speech processed to preserve only the amplitude envelopes of the original signal do not preserve that kind of dynamic spectral structure particularly well, whereas sine wave replicas do. Consequently, how well listeners of different ages can perceptually organize various kinds of signals seems partly related to the signal properties themselves. If the processed signal preserves properties to which listeners are especially sensitive, those listeners are better able to recover the auditory (or in the case of speech, the linguistic) object.

The current study was conducted to extend Nittrouer *et al.* (2009) by examining how children younger than 7 years of age, the youngest group in that study, perceptually organize processed speech signals. Because children younger than 7 years of age have consistently demonstrated even stronger weighting of dynamic spectral components in the speech signal, they should show even a stronger advantage in their recognition of the sine wave replicas over the amplitude envelope signals. However, the 7-year-olds in that study recognized fewer than 30% of the words correctly in two out of three conditions, and so we worried that younger children might not recognize any words at all if the same or similar stimuli were used. Therefore, we elected to use sentences that provided both syntactic and semantic constraints. Children as young as 3 years of age participated in this experiment. The goal was to examine listeners' abilities to perceptually organize the sensory information provided by dynamic spectral structure (as preserved by sine waves) and by amplitude envelopes. Four-channel amplitude envelope signals were used because adults had shown identical performance with those signals to that obtained with sine wave speech in Nittrouer *et al.* (2009). Native Mandarin-speaking adults participated again because their inclusion could help bolster arguments that age-related differences in speech recognition, if found, are likely not due to differences in auditory processing of the sensory information.

Our focus in this experiment was on the perceptual organization of linguistically relevant acoustic structure in the speech signal. Because we used sentence materials, however, we needed a way to ensure that any group differences observed were not due to linguistic context effects. Miller *et al.* (1951) are generally credited with being the first investigators to examine differences in speech recognition scores due to the context, in which the stimulus is presented. They showed that when isolated words presented in background noise were part of small sets made known to research participants, recognition was more accurate than when set size was larger. As their metric they used the difference in signal-to-noise ratio (SNR) needed to achieve a preselected target of 50% correct recognition. It was not long before other investigators extended the notion of set size to frequency of word usage: Commonly occurring words can be recognized at poorer SNRs than uncommonly occurring words (e.g., Broadbent, 1967; Howes, 1957; Rosenzweig and Postman, 1957). In the rhetoric of information theory, word frequency influences the number of effective channels of sensory information available to the perceiver.

Boothroyd (1968) developed several metrics that provide a way of quantifying the numbers of effective channels of information used by the perceiver. The metric of most relevance to the current work is known as the *j* factor and derives from the fact that the probability of recognizing a whole word is dependent on the probabilities of recognizing the separate parts (or phonemes) forming that word. If lexicality played no role in recognition, then the probability of recognizing a whole word correctly would be directly related to the probability of recognizing each of the parts, such that

$$p_w = p_p^n, \tag{1}$$

where p_w is the probability of recognizing the whole word, p_p is the probability of recognizing each part, or phoneme, and n is the number of parts in the whole. However, this relation holds only when it is necessary to recognize each part in order to recognize the whole, as happens with nonwords, for instance. When listening to real words, the lexical status of those words generally reduces the need to recognize each separate part in order to recover the whole. Therefore, we can change n to a dimensionless exponent, such as j, to quantify the effect of hearing these phonemes in a lexical context. This exponent varies between 1 and n, and the extent to which it is smaller than n indicates the magnitude of the context effect. Now Eq. (1) can be changed to

$$p_w = p_p^j, \tag{2}$$

where j is the number of independent channels of information, and we can solve for the effective number of channels of information in the word with

$$j = \log(p_w)/\log(p_p). \tag{3}$$

Boothroyd and Nittrouer (1988) demonstrated the validity of the j factor with the finding that j equaled 3.07 when listeners were asked to recognize nonsense CVCs. That value was not statistically different from the value of 3, which would be expected if listeners need to recognize each segment independently in order to recognize the whole syllable. By contrast, Boothroyd and Nittrouer (1988) reported a mean j of 2.46 when these same listeners were asked to recognize CVC real words. That value was statistically different from the 3.07 obtained for nonsense syllables. Together, these findings support the use of the j factor as a metric of the number of independent channels of information needed for recognition. This metric has been used by others for several clinical purposes such as assessing the abilities of deaf children to use lexical context (Boothroyd, 1985). Benkí (2003) used the j factor to index the lexical context effect of high-frequency words compared to low-frequency words. In addition to demonstrating a stronger context effect for high- over low-frequency words (j's of 2.25 and 2.46, respectively), Benkí (2003) replicated Boothroyd and Nittrouer's finding of a j equal to 3.07 for nonsense syllables.

Sentence context similarly improves listeners' abilities to recognize words under conditions of signal degradation (e.g., Boothroyd and Nittrouer, 1988; Duffy and Giolas,

1974; Giolas *et al.*, 1970; Kalikow *et al.*, 1977; Nittrouer and Boothroyd, 1990). When examining the effects of sentence context on word recognition, investigators are generally not interested in the effect of lexicality, but rather in the effects of syntactic and semantic constraints. Those effects can be quantified using Eq. (3) by making p_w the probability of recognizing the whole sentence and p_p the probability of recognizing each word in the sentence.

Nittrouer and Boothroyd (1990) computed the number of independent channels of information needed by adults and children (4–6 years of age) listening to sentences in their native language with signals degraded by the addition of noise at two SNRs: −3 dB and 0 dB. Two kinds of sentences were used: sentences that had appropriate syntactic structure, but no useful semantic information (syntax-only sentences), and sentences that provided clear syntactic structure, as well as strong semantic constraints (syntax+semantics sentences). All were comprised of four monosyllabic words. Across the two SNRs employed in that study, the mean *j* for adults listening to the syntax+semantics sentences was 2.32, and the mean *j* for children listening to those same sentences was 2.59. This age-related difference was not statistically significant, and so it may be concluded that children as young as 4 years are able to use syntactic structure and semantic context as well as adults, at least for simple sentences.

The *j* factor has not been used to assess the contributions of sentence context to recognition by adults listening to a second language, but several investigators have used the speech in noise (SPIN) test to evaluate this effect. In the SPIN test, listeners hear two kinds of sentences and must report the last word in each. In one kind of sentence, the words prior to the last word provide semantic information that strongly predicts that word (e.g., *My clock was wrong so I got to school late; After my bath I dried off with a towel.*). In the other kind of sentence, the preceding words provide no information that would especially predict the final word (e.g., *Mom talked about the pie. He read about the flood.*). Mayo *et al.* (1997) presented these sentences at several SNRs to L1 English listeners and L2 listeners who had learned English at ages varying from birth to early adulthood. Results showed that listeners who had learned English after the age of 14 years were less capable than native and early L2 learners at using the highly predictable sentence context to aid recognition of the final word. In another study, Bradlow and Alexander (2007) presented these same sentences to native and L2 English listeners, but used samples that could be described as plain or clear speech. Listeners in both groups showed more accurate recognition for words in the high-predictability contexts, but for L2 listeners this advantage was restricted to clear sentences.

The results above suggest that L2 adults are poorer at using sentence context for speech recognition than L1 adults. However, those results were obtained using the SPIN test, and there are important differences between that task and the way *j* factors are computed. In the SPIN test, recognition of only words in sentence-final position is being evaluated, and the comparison is between sentences that have strong semantic constraints (high predictability) and those that have neutral semantic constraints (low predictability). With *j* factors, recognition scores for all words within the sentences are involved in computation, and every word serves as context for every other word. The difference between these two methods is highlighted by the fact that young children show diminished semantic context effects on the SPIN test compared to older children (Elliott, 1979), but young children have *j* factors equivalent to those of adults, as long as syntax and semantics are within their language competencies (Nittrouer and Boothroyd, 1990). On the SPIN test, sentences with high and low predictabilities tend to differ in syntactic structure, with high-predictability sentences generally having more complex structures. That could constrain the abilities of children and L2 listeners to use the sentence context if the syntactic constructions are not within their competencies. This is not a concern when *j* factors are computed, both because the sentences that have been used involve simple syntactic constructions and because wholes and parts are scored from the same responses. Because of these differences, we did not necessarily expect L2 listeners to show the same deficits in using sentence context compared to L1 listeners that have been reported earlier.

In summary, the current experiment was designed to examine how listeners of different ages and native-language experience perceptually organize two kinds of signals derived from speech: sine wave replicas and amplitude envelopes. We hypothesized that L2 adults would be equally poor at recognition with both kinds of signals, but that children would show an advantage for the sine wave replicas over the amplitude envelope signals because they heavily weight global spectral structure in their speech perception. We selected sentence materials that we thought would produce similar context effects across listeners, but planned to compute *j* factors as a test of that assumption.

II. METHOD

A. Participants

100 listeners participated in this experiment: 80 native speakers of American English (20 adults, 20 7-year-olds, 20 5-year-olds, and 20 3-year-olds) and 20 adult, native speakers of Mandarin Chinese who were competent L2 speakers of English. None of the participants had listened to sine wave replicas or amplitude envelope speech before. All L2 speakers had started learning English between the ages of 12 and 14 years in language classes in China. They all came to the United States between the ages of 21 and 26 years to study, and were in the United States as graduate students, postdoctoral fellows, or junior faculty members. Mean age of the L1 adults was 24 years (range 19–35 years), and mean age of the L2 adults was 26 years (range 22–36 years). To participate, all listeners had to pass a hearing screening and demonstrate appropriate abilities to comprehend and produce spoken English. 3-year-olds were also given the Goldman–Fristoe 2 test of articulation (Goldman and Fristoe, 2000) to screen for speech problems. They were required to score at or better than the 30th percentile for their age. Each group was comprised of 10 males and 10 females, except for 7-year-olds (9 females, 11 males) and L2 listeners (11 females, 9 males).

B. Equipment

All speech samples were recorded in a sound booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Laboratories Soundblaster analog-to-digital converter. Perceptual testing took place in a sound booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Samson headphone amplifier, and AKG-K141 headphones.

C. Stimuli

Seventy-two sentences were used as stimuli in this experiment: 12 for practice and 60 for testing. All were selected from the hearing in noise test for children (HINT-C) (Nilsson *et al.*, 1996; Nilsson *et al.*, 1994), and all were five words in length. The sentences selected for use are shown in the appendix. The HINT-C sentences were originally developed to test how well children can recognize speech in competing noise. They are five to eight words long, syntactically correct, and follow a subject-predicate structure. They are highly predictable semantically. Sentences from the HINT-C were used by Eisenberg *et al.* (2000) in an experiment looking at recognition of amplitude envelopes by adults and children. Those listeners, regardless of age, had close to 90% correct word recognition when listening to eight-channel envelope stimuli. With four-channel stimuli, adults had close to 70% correct word recognition and 5- to 7-year-olds had close to 35% correct word recognition. Thus, although our primary motivation for using four channels for the amplitude envelope stimuli was because Nittrouer *et al.* (2009) found that adults had equivalent word recognition for those stimuli and sine wave replicas, the Eisenberg *et al.* (2000) results supported the expectation that word recognition scores would be neither too close to 0 nor too close to 100%, at least for the amplitude envelope stimuli. Finally, we restricted our corpus to five-word sentences so that length would be consistent, and relatively brief, thereby minimizing concerns about memory.

The selected sentences were recorded at a 44.1-kHz sampling rate with 16-bit digitization by an adult male speaker of American English who is a trained phonetician. All sentences were equalized for mean rms amplitude across sentences before any processing was done, and all sentences were used to create sine wave (SW) and four-channel envelope signals. Because these latter stimuli preserve the amplitude envelopes in each of four channels, they are dubbed as AE stimuli in this report. For creating the SW stimuli, a PRAAT routine written by Darwin (http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS) was used to extract the center frequencies of the first three formants. Stimuli were generated from these formant frequencies, and spectrograms were compared to spectrograms of the original sentences to ensure that the trajectories of the sine waves matched those of the formant frequencies.

To create the AE stimuli, a MATLAB routine was written. All signals were first low-pass filtered with an upper cut-off frequency of 8000 Hz. For the four-channel stimuli, cut-off frequencies between bands were 800, 1600, and 3200 Hz. Each channel was half-wave rectified, and results used to modulate white noise, limited by the same band-pass filters as those used to divide the speech signal into channels. Resulting bands of modulated noise were low-pass filtered using a 160-Hz high-frequency cut-off, and combined.

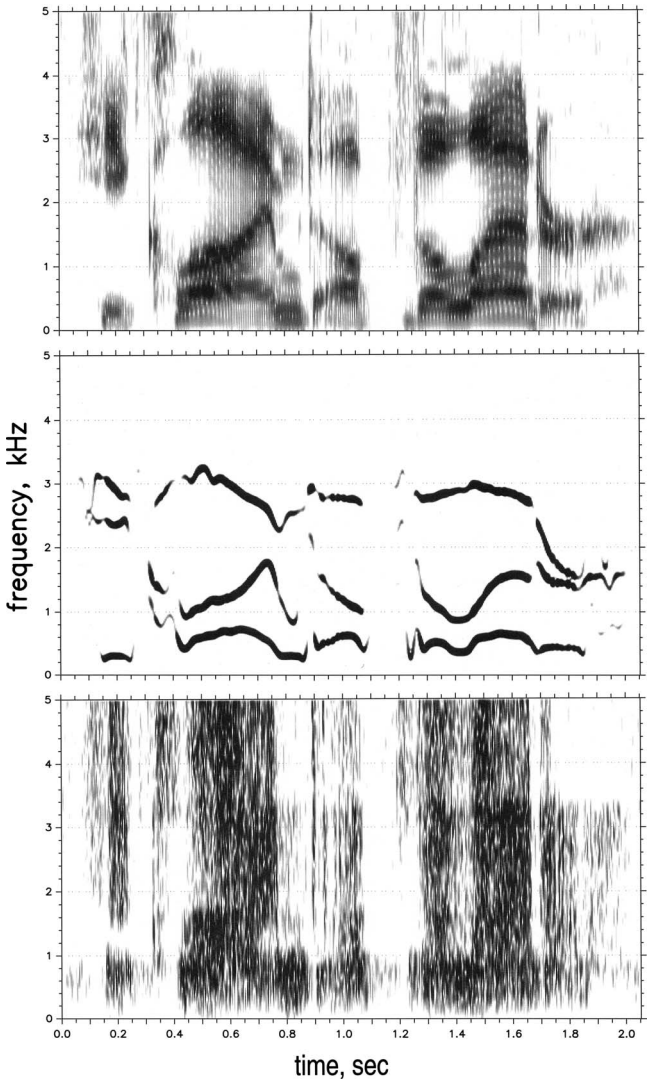


FIG. 1. Spectrograms of the sentence "He climbed up the ladder." (Top) Natural speech sample from which SW and AE stimuli were derived. (Middle) SW stimulus. (Bottom) Four-channel AE stimulus.

Figure 1 displays a natural speech sample of *He climbed up the ladder* (top panel), the SW replica (center panel), and the AE replica (bottom panel).

D. Procedures

All stimuli were presented under headphones at a peak intensity of 68 dB sound pressure level. For each participant, the software randomly selected 30 sentences to present as AE and 30 to present as SW prior to testing. Half of the participants heard all 30 AE sentences first, and then the SW sentences, and half of the participants heard all 30 SW sentences first, and then the AE sentences. Training was the same for each condition. For each of the six practice sentences, the

TABLE II. Mean percent correct words recognized by each group in each condition. Standard deviations are in parentheses. AE refers to amplitude envelope stimuli and SW refers to sine wave stimuli.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	79.53 (8.30)	43.85 (16.63)	30.73 (17.93)	16.23 (11.53)	33.47 (12.29)
SW	98.37 (1.38)	91.90 (3.03)	86.32 (6.11)	75.40 (15.22)	77.20 (8.78)

natural version was played first. The listener was instructed to listen and repeat it. Next, the listener was told "Now you will hear the sentence in a robot voice. Listen carefully so you can repeat it." The notion of a robot was invoked to personify the signal so that listeners would more readily hear the stimuli as speech. The children were specifically told that the SW sentences were spoken by a robot with a "squeaky voice," and the AE sentences were spoken by a robot with a "scratchy voice." For both the AE and SW conditions, listeners were asked at the end of the six-sentence practice if they were able to recognize the "robot's" productions as speech. It was made clear that the question was whether they could hear the stimuli as speech, rather than could they understand every word. If a listener either was unable to repeat any of the unprocessed sentences or failed to report hearing the processed signals as speech, he was dismissed. None of the listeners had difficulty hearing the AE sentences as speech. Two additional participants, one 3-year-old and one 5-year-old, were dismissed because they did not report hearing the SW sentences as speech.

During testing, the order of presentation of the sentences was randomized independently for each listener. Each sentence was played once, and the listener repeated it as best as possible. Participants could ask for a sentence to be replayed one time. Nittrouer *et al.* (2009) used three trials of each sentence, but listeners in all groups showed similar improvements across repetitions and means from the first to the third repetition never differed by more than 1 standard deviation. Consequently, the outcomes of that study would not have been different if only one repetition had been used, and so we had no motivation to use more repetitions in this study. The number of incorrect words for each sentence was entered into the program interface during testing. 5- and 7-year-olds moved a game piece along a game board every ten sentences, while 3-year-olds used a stamp to mark a paper grid after every five sentences, in order to maintain attention and give a sense of how much of the task remained.

After hearing the sentences in their processed forms, all sentences were played to listeners in their unprocessed forms. Listeners could get no more than 10% of the words wrong on this task. Data from two additional participants, one L2 adult and one 3-year-old, were excluded from analyses because they made too many errors repeating the unprocessed sentences. None of the other listeners had trouble comprehending the unprocessed sentences.

In addition to entering the number of words repeated incorrectly at the time of testing, participants were audiorecorded. A graduate student later listened to 15% of the recordings, and marked the number of words she heard as incorrect. Reliability coefficients were 1.00 for both AE and SW sentences for listeners of all ages.

III. RESULTS

A. Word recognition

The percentage of words correctly recognized across all sentences within any one condition served as the dependent measure. All statistics were performed using arcsine transforms because some listeners scored above 90% correct in the SW condition and others scored below 10% correct in the AE condition. As a preliminary test, analyses of variance (ANOVAs) were conducted on percent correct recognition scores for each group separately, based on whether they heard AE or SW sentences first. No significant order effect was observed for any group.

Next, the effect of English experience for the L2 adults was examined. Pearson product-moment correlation coefficients were computed between recognition scores for both the AE and SW sentences and the length of time the L2 adults had been speaking English based on when they began language classes. The correlation coefficient for AE sentences was -0.04, and for SW sentences it was -0.13. Neither was significant. In addition, correlation coefficients were computed between the numbers of years these participants had lived in the United States and their recognition scores. In this case, the correlation coefficient for AE sentences was -0.41, and for SW sentences, it was -0.27. Again, neither of these was significant; in fact, slightly negative correlations were found. The lack of an L2 experience effect matches findings by others: If learning starts after puberty, eventual proficiency is variable and unrelated to factors such as age of initial exposure or age of arrival in the new country (e.g., Johnson and Newport, 1989).

Table II shows mean correct word recognition for each group for each kind of stimulus. The greatest differences across groups occurred in the AE condition, with a difference of roughly 63% between the highest scoring group (L1 adults) and the lowest scoring group (3-year-olds). Groups performed more similarly in the SW condition, with a difference of only 23% between the highest and lowest scoring groups (again, L1 adults and 3-year-olds). This difference across conditions was due to 3-year-olds performing much better in the SW than in the AE condition, rather than to a decrement in performance for L1 adults in the AE condition. In fact, adults performed better in the SW than in the AE condition, just less so than children. When a two-way ANOVA was performed on these scores, with group as the between-subjects factor and signal type as the within-subjects factor, both main effects were significant: group, $F(4,95)=68.93$, $p<0.001$, and signal types, $F(1,95)=995.93$, $p<0.001$. Of particular interest, the group

TABLE III. Results of the statistical analyses of percent correct scores for group effects, using arc sine transforms. *F* ratios (shown in bold) are for one-way ANOVAs, with group as the between-subjects factor; t-tests are for *posthoc* comparisons. Degrees of freedom are 4, 95 for the ANOVAs, and 95 for the t-tests. Computed *p* values are shown. Bonferroni corrections for five multiple comparisons require a *p* of less than 0.005 for the comparison to be significant at the 0.05 level. Comparisons meeting this adjusted level are indicated by an asterisk.

	F ratio or *t*	*p*
AE sentences	**49.42**	<0.**001**
L1 adults vs 7-year-olds	7.22	<0.001*
L1 adults vs 5-year-olds	10.16	<0.001*
L1 adults vs 3-year-olds	13.31	<0.001*
L1 adults vs L2 adults	9.17	<0.001*
7- vs 5-year-olds	2.94	0.004*
7- vs 3-year-olds	6.09	<0.001*
7- vs L2 adults	1.95	0.054
5- vs 3-year-olds	3.16	0.002*
5- vs L2 adults	−0.99	NS
3- vs L2 adults	−4.14	<0.001*
SW sentences	**46.24**	<0.**001**
L1 adults vs 7-year-olds	5.03	<0.001*
L1 adults vs 5-year-olds	7.64	<0.001*
L1 adults vs 3-year-olds	11.60	<0.001*
L1 adults vs L2 adults	11.25	<0.001*
7- vs 5-year-olds	2.62	0.010
7- vs 3-year-olds	6.57	<0.001*
7- vs L2 adults	6.23	<0.001*
5- vs 3-year-olds	3.95	<0.001*
5- vs L2 adults	3.61	<0.001*
3- vs L2 adults	−0.34	NS

×signal type interaction was significant, $F(4,95)=12.64$, $p<0.001$, indicating that the variability in magnitude of the signal effect across groups was significant.

Table III shows the results of one-way ANOVAs and *posthoc* comparisons performed on word recognition scores for AE and SW stimuli separately. Significant main effects of group were found for both the AE and SW conditions, so *posthoc* comparisons among groups were examined. For the AE condition, L1 adults performed best, and mean correct recognition differed significantly from every other group. Differences were observed among all children's groups, with 7-year-olds performing the best, followed by 5-year-olds, and then 3-year-olds. Although 7-year-olds had better recognition scores than L2 adults, this difference was not significant once a Bonferroni correction was applied. However, this is the one statistical outcome that differed for untransformed and transformed data: When the ANOVA and *posthoc* comparisons were performed on untransformed scores, this comparison was unambiguously significant, $t(95)=2.38$, $p=0.019$, supporting the assertion that 7-year-olds performed more accurately than L2 adults. Five-year-olds performed similarly to L2 adults, but 3-year-olds performed significantly more poorly.

For the SW stimuli, L1 adults again performed the best, with a mean recognition score that differed significantly from every other group. 7- and 5-year-olds performed similarly, and the difference between their mean scores was not significant when a Bonferroni correction was applied. Both groups performed significantly better than 3-year-olds and L2 adults. 3-year-olds did not differ from L2 adults.

Prior to running this experiment, there was no reason to expect adults to be more accurate in their responses for one condition over the other: AE stimuli can be created with any number of channels, and we had selected four channels precisely because adults in Nittrouer *et al.* (2009) showed identical performance with four-channel AE and SW stimuli. 7-year-olds in that experiment were the only listener group to show better recognition for SW than for AE stimuli, and the difference between the two conditions was 12 percentage points. For this experiment, Table II indicates that, on the whole, listeners performed better with SW sentences than with AE sentences. That general result makes it important to investigate the magnitude of this condition-related difference across listener groups, and Table IV shows means of individual difference scores for each group. The L1 adults demonstrated the smallest difference across conditions, and the youngest children demonstrated the greatest difference. It may be that difference scores for L1 adults were constrained because many of these listeners showed near-perfect accuracy, especially with SW stimuli, but there can be no question that 3-year-olds showed an advantage for SW over AE stimuli. A one-way ANOVA done on these difference scores with group as the between-subjects factor revealed a significant effect, $F(4,95)=25.24$, $p<0.001$. *Posthoc* comparisons showed significant differences between L1 adults and all other groups. The comparison of 7- and 3-year-olds was significant. In addition, L2 adults showed significantly smaller difference scores than either 5- or 3-year-olds. Because the scores of these adults were not constrained by ceiling or floor effects, the finding suggests that the age-related difference for L1 listeners in condition effects is likely not entirely due to L1 adults scoring near the ceiling. Children were disproportionately more advantaged for the SW than the AE stimuli.

B. Top-down effects

Using the formula presented in the Introduction, *j* factors were computed for individual listeners using word and sentence recognition scores. Because this computation requires the use of sentence recognition scores, mean percentages of sentences recognized correctly for each group are shown in Table V, for AE and SW stimuli separately. Table VI shows mean *j* factors for each group. Scores for indi-

TABLE IV. Mean difference scores across the SW and AE conditions. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
Difference	18.83 (8.42)	48.05 (15.58)	55.58 (19.84)	59.17 (13.64)	43.73 (10.25)

TABLE V. Mean percent correct sentences recognized by each group in each condition. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	61.00 (12.66)	19.00 (10.93)	8.50 (8.89)	0.67 (1.37)	7.50 (6.66)
SW	93.50 (4.90)	78.50 (7.45)	63.83 (10.88)	44.33 (18.26)	43.65 (13.91)

vidual listeners were not included if the number of words or sentences recognized correctly was greater than 95% or less than 5%. For the SW sentences, 19 out of 20 L1 adults had better than 95% correct word recognition. For the AE sentences, all 20 3-year-olds had poorer than 5% correct sentence recognition.

One-way ANOVAs were computed on individual *j* factors, for the AE and the SW stimuli separately. For the SW stimuli, scores for L1 adults were excluded. For the AE stimuli, scores for the 3-year-olds were excluded. Neither analysis revealed a significant *F* ratio, so it may be concluded that listeners in all groups had similar context effects.

Table VII shows mean *j* factors only for participants who had word and sentence recognition scores between 5% and 95% correct in both conditions. A two-way ANOVA was computed on these scores, with group as the between-subjects factor and stimulus type as the within-subjects factor. The effect of stimulus type was significant, $F(1,33)=43.27$, $p<0.001$, but not the group effect. The stimulus type × group interaction was not significant. Overall then, these *j* factors were smaller for the AE sentences than for the SW sentences, and that difference was consistent across listener groups.

C. Perceptual organization or top-down effects

The main findings emerging from these data are that the younger children showed an advantage in their perceptual organization of the SW stimuli over the AE stimuli, and the L2 adults were poor at perceptually organizing either type of signal. However, two constraints in the results could dampen enthusiasm for those conclusions: First, L1 adults showed ceiling effects for the SW stimuli. Consequently, the true difference in scores between the SW and AE conditions may not have been obtained for that group. The second constraint was that *j* factors could not be computed for all listeners. Possibly the listeners for whom *j* factors could not be computed accounted for the observed differences in signal effects across groups.

Those concerns can be ameliorated by looking at results for only those participants who provided *j* factors. Mean correct word recognition for these individuals are shown in Table VIII, and when these results are compared to those for all listeners shown on Table II, it is apparent that the same trends across groups are observed. Analyses of covariance performed on these values, using *j* factor as the covariate, resulted in significant group effects, for both the AE stimuli, $F(3,54)=75.88$, $p<0.001$, and the SW stimuli, $F(3,72)=18.89$, $p<0.001$.

Further support for the finding that young children did indeed show an advantage in their perceptual organization of SW stimuli, compared to adults, is obtained by looking across results for 5-year-olds and L2 adults. Table III indicates that listeners in these two groups scored the same with the AE stimuli. Tables VI and VII show that top-down effects were similar for listeners in the two groups. The only result that differs for listeners in the two groups concerns their performance with the SW stimuli: 5-year-olds scored significantly better than L2 adults, as indicated by a significant *posthoc* comparison (Table III). Even when scores are examined for only those 5-year-olds and L2 adults for whom *j* factors could be computed (Table VIII), it is apparent that listeners in both groups performed similarly for the AE stimuli, but 5-year-olds scored better with the SW stimuli.

Finally, it is important to the conclusions reached here that context effects were less for the signal condition in which listeners scored better: that is, the SW condition. That outcome indicates that it was not simply the case that something about the SW stimuli permitted stronger context effects. Instead, the enhanced recognition that young children showed for the SW stimuli seem to be due to their abilities to perceptually organize these signals more handily.

IV. DISCUSSION

The experiment reported here was conducted largely to examine whether the way in which listeners perceptually organize speech signals depends on their native language experience, and on the amount of that experience. An additional question concerned whether listeners are better able to recover relevant linguistic structure with some kinds of processed signals than with others: In particular, we asked whether listeners are better able to recover structure from signals that preserve sensory information to which they are especially sensitive.

TABLE VI. Mean *j* factors for the AE and SW conditions. N/A signifies that participants in that group either all scored above 95% correct (adults, SW words) or below 5% correct (3-year-olds, AE sentences). Standard deviations are in parentheses. The numbers of participants with *j* scores for each group are given in italics.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	2.26 (0.34), *20*	2.26 (0.55), *18*	2.43 (0.60), *10*	N/A	2.43 (0.48), *11*
SW	N/A	2.98 (0.94), *17*	3.36 (0.92), *20*	3.43 (0.82), *20*	3.42 (0.53), *20*

TABLE VII. Mean *j* factors for the AE and SW conditions for participants who had *j* scores for both conditions. The numbers of participants are given in italics under the group name. Standard deviations are in parentheses.

	7-year-olds (*15*)	5-year-olds (*10*)	L2 adults (*11*)
AE	2.23 (0.59)	2.43 (0.60)	2.43 (0.48)
SW	3.03 (0.99)	3.65 (0.81)	3.49 (0.67)

Looking first at results for adults, it was found that the non-native listeners had poorer speech recognition for both kinds of signals than did the native listeners. All participants were required to pass a hearing screening before testing, and so all had equal access to the sensory information available in these signals. Furthermore, *j* factors were similar for both groups of adults, so it may be concluded that top-down cognitive influences on speech recognition were similar in this experiment: Although others have shown effects for L2 adults, none were found for these sentences with simple syntax. Consequently, we are left to posit the locus of effect squarely on how the two groups perceptually organized the sensory information that they received. The suggestion made here is that every language has its own unique global structure. The long-term spectra computed by Boysson-Bardies *et al.* (1986) for samples from speakers of four different languages demonstrated this fact for spectral structure. Presumably languages also differ in terms of global temporal and amplitude structure, given that languages vary in rhythm and stress. Here, it is suggested that adults are familiar with the global structure of their native language, and organize speech signals accordingly. This idea is commensurate with what is known about the phonetic perception of listeners for non-native languages. Listeners organize sensory information in the speech signal according to the phonetic structure of their native language (Beddor and Strange, 1982; Abramson and Lisker, 1967; MacKain *et al.*, 1981), so much so, in fact, that the effect is often described as L1 interference (e.g., Flege and Port, 1981). However, ideas regarding how L1 interferes with the perception of L2 traditionally focus on sensory information at the level of the acoustic cue. The signals presented to listeners in this experiment lacked that level of signal detail; there were no properties constituting acoustic cues. Nonetheless, these L2 speakers had more difficulty than L1 listeners recognizing linguistically relevant structure. That can only mean that they were unable to perceptually organize the sensory information in the signals in such a way as to recover global structure in the L2.

The children in this study were obviously not affected by having a native language interfere with how they perceptually organized the signals they heard because they were listening to processed versions of their native language. Nonetheless, these children had poorer speech recognition for the processed signals than L1 adults. This age-related finding likely reflects the fact that children have not had ample time to become as familiar with the global structure of their native language as adult listeners. Of course, it could just be that children show deficits at recognizing any impoverished signal, but results of another study suggest otherwise. Nittrouer and Lowenstein (2009) presented fricative-vowel syllables to adults and children for fricative labeling. The fricative noises were synthetic and spanned a continuum from one appropriate for /s/ to one appropriate for /ʃ/. The vocalic portions were either from natural speech samples or were sine wave replicas of that natural speech. Children's labeling functions were indistinguishable across the two kinds of stimuli. Therefore, it may be concluded that children are capable of using the information in processed speech signals for recognition, at least at the level of phonetic labeling.

Alternatively, the age-related differences observed here could be attributable to children having poorer access to the sensory information available in the speech signal, perhaps because of immature auditory systems (e.g., Sussman, 1993). That seems unlikely, however, given that L2 adults also performed more poorly than L1 adults, and no one would attribute that difference to poorer access to sensory information.

Moreover, it was found that children had similar *j* factors to adults, and so age-related differences in word recognition cannot be attributed to differences in top-down effects. Again, it seems that the source of the variability in recognition scores for L1 listeners of different ages must be placed squarely on differences in abilities to perceptually organize the sensory information in order to recover linguistically relevant structure.

Finally, it is informative to examine relative performance for the two kinds of processed signals used in this experiment. Children showed disproportionately better performance for the SW stimuli than for the AE stimuli. For example, 5-year-olds performed similarly to L2 adults for AE stimuli but significantly better for SW stimuli. 3-year-olds, on the other hand, performed more poorly than L2 adults for AE stimuli, but similarly with SW stimuli. Thus, all listeners were better able to recognize speech with the SW stimuli than for the AE stimuli, but this advantage was disproportionately greater for the younger children. We speculate that the reason for this outcome is related to the fact that in recognizing speech, young children rely particularly strongly on the dynamic spectral components, which are preserved by sine wave replicas.

In conclusion, this experiment was designed as an extension of Nittrouer *et al.* (2009), which showed that native language experience affected listeners' abilities to perceptu-

TABLE VIII. Mean percent correct words recognized by listeners for whom *j* factors could be computed. Standard deviations are in parentheses.

	L1 adults	7-year-olds	5-year-olds	3-year-olds	L2 adults
AE	79.53 (8.30)	47.35 (13.37)	43.20 (10.08)	N/A	40.97 (9.78)
SW	N/A	91.14 (2.59)	86.32 (6.11)	75.40 (15.22)	77.20 (8.78)

ally organize processed speech signals, and that 7-year-olds showed greater benefits for SW over AE stimuli than adults. First, this experiment replicated the effect of native language experience: L2 adults performed more poorly than L1 listeners, in spite of having sufficient skills in their L2 to take graduate level course work. This difference in performance could not be attributed either to differences in sensory functioning or to differences in top-down cognitive effects. Second, this experiment extended the age-related findings of Nittrouer *et al.* (2009): Younger children than those who participated in that earlier study showed disproportionately enhanced performance for SW stimuli. 3-year-olds were able to recognize 75% of the words presented in SW sentences correctly. In sum, both native language experience and the extent of that experience have been found to influence listeners' abilities to organize sensory information in order to recover linguistically relevant structure.

These findings could have important implications for how we intervene with individuals, particularly children, who have hearing loss. At present, cochlear implants are not very good at preserving the kind of dynamic spectral information represented so well by the sine wave replicas used here. However, some individuals with severe-to-profound hearing loss have aidable hearing in the low frequencies. It may be worth exploiting that residual hearing in order to provide at least the first formant to these individuals through a hearing aid, and this practice could be especially beneficial for young children who rely heavily on dynamic spectral structure. Uchanski *et al.* (2009) found evidence to support this recommendation from one child with electric and acoustic hearing on the same side, and Nittrouer and Chapman (2009) found evidence from language outcomes of 58 children with hearing loss: Those who had some experience with combined electric and acoustic hearing fared better than children who had no such experience.

ACKNOWLEDGMENTS

We thank Mallory Monjot for her help with the reliability measures. This work was supported under research Grant No. R01 DC-00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to S.N.

APPENDIX

Sentences from the HINT-C that were used are as follows. The words in parentheses indicate that either words are acceptable answers. The underlined word is the one that is said in creating the stimuli.

(a) Practice sentences

1. The yellow pears taste good.
2. (A/the) front yard (is/was) pretty.
3. The pond water (is/was) dirty.
4. The ground (is/was) very hard.
5. They painted (a/the) wall white.
6. (A/the) little girl (is/was) shouting.
7. (An/the) ice cream (is/was) melting.
8. (An/the) apple pie (is/was) baking.
9. (A/the) boy forgot his book.
10. The two farmers (are/were) talking.
11. The sky (is/was) very blue.
12. (A/the) black dog (is/was) hungry.

(b) Test sentences

1. Flowers grow in (a/the) garden.
2. She looked in her mirror.
3. They heard (a/the) funny noise.
4. (A/the) book tells (a/the) story.
5. (A/the) team (is/was) playing well.
6. (A/the) lady packed her bag.
7. They waited for an hour.
8. (A/the) silly boy (is/was) hiding.
9. (A/the) mailman shut (a/the) gate.
10. (A/the) dinner plate (is/was) hot.
11. They knocked on (a/the) window.
12. He (is/was) sucking his thumb.
13. He grew lots of vegetables.
14. He hung up his raincoat.
15. The police helped (a/the) driver.
16. He really scared his sister.
17. He found his brother hiding.
18. She lost her credit card.
19. He wore his yellow shirt.
20. The young people (are/were) dancing.
21. Her husband brought some flowers.
22. The children washed the plates.
23. (A/the) baby broke his cup.
24. They (are/were) coming for dinner.
25. They had a wonderful day.
26. The bananas (are/were) too ripe.
27. She argues with her sister.
28. (A/the) kitchen window (is/was) clean.
29. (A/the) mother heard (a/the) baby.
30. (An/the) apple pie (is/was) good.
31. New neighbors (are/were) moving in.
32. (A/the) woman cleaned her house.
33. The old gloves (are/were) dirty.
34. (A/the) painter uses (a/the) brush.
35. The bath water (is/was) warm.
36. Milk comes in (a/the) carton.
37. (A/the) ball bounced very high.
38. School got out early today.
39. The rain came pouring down.
40. (A/the) train (is/was) moving fast.
41. (A/the) baby slept all night.
42. Someone (is/was) crossing (a/the) road.
43. (A/the) big fish got away.
44. (A/the) man called the police.
45. (A/the) mailman brought (a/the) letter.
46. He climbed up (a/the) ladder.
47. He (is/was) washing his car.
48. The sun melted the snow.
49. The scissors (are/were) very sharp.
50. Swimmers can hold their breath.
51. (A/the) boy (is/was) running away.
52. (A/the) driver started (a/the) car.
53. The children helped their teacher.
54. (A/the) chicken laid some eggs.
55. (A/the) ball broke (a/the) window.
56. Snow falls in the winter.
57. (A/the) baby wants his bottle.
58. (An/the) orange (is/was) very sweet.

59. (An/the) oven door (is/was) open.
60. (A/the) family bought (a/the) house.

Abramson, A. S., and Lisker, L. (**1967**). "Discriminability along the voicing continuum: Cross-language tests," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, pp. 569–573.

Beddor, P. S., and Strange, W. (**1982**). "Cross-language study of perception of the oral-nasal distinction," J. Acoust. Soc. Am. **71**, 1551–1561.

Benkí, J. R. (**2003**). "Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition," J. Acoust. Soc. Am. **113**, 1689–1705.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (**1989**). "Discovering phonetic coherence in acoustic patterns," Percept. Psychophys. **45**, 237–250.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (**2004**). "Neural correlates of sensory and decision processes in auditory object identification," Nat. Neurosci. **7**, 295–301.

Boothroyd, A. (**1968**). "Statistical theory of the speech discrimination score," J. Acoust. Soc. Am. **43**, 362–367.

Boothroyd, A. (**1985**). "Evaluation of speech production of the hearing impaired: Some benefits of forced-choice testing," J. Speech Hear. Res. **28**, 185–196.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Boysson-Bardies, B., de Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigations of cross-linguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterstrom, (Stockton, New York), pp. 113–126.

Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Broadbent, D. E. (**1967**). "Word-frequency effect and response bias," Psychol. Rev. **74**, 1–15.

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.

Dau, T., Ewert, S., and Oxenham, A. J. (**2009**). "Auditory stream formation affects comodulation masking release retroactively," J. Acoust. Soc. Am. **125**, 2182–2188.

Dorman, M. F., Lindholm, J. M., and Hannley, M. T. (**1985**). "Influence of the first formant on the recognition of voiced stop consonants by hearing-impaired listeners," J. Speech Hear. Res. **28**, 377–380.

Duffy, J. R., and Giolas, T. G. (**1974**). "Sentence intelligibility as a function of key word selection," J. Speech Hear. Res. **17**, 631–637.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

Elliott, L. L. (**1979**). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Erber, N. P. (**1971**). "Evaluation of special hearing aids for deaf children," J. Speech Hear Disord. **36**, 527–537.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Flege, J. E., Schmidt, A. M., and Wharton, G. (**1996**). "Age of learning affects rate-dependent processing of stops in a second language," Phonetica **53**, 143–161.

Fu, Q., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Giolas, T. G., Cooker, H. S., and Duffy, J. R. (**1970**). "The predictability of words in sentences," J. Aud Res. **10**, 328–334.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).

Gottfried, T. L., and Beddor, P. S. (**1988**). "Perception of temporal and spectral information in French vowels," Lang Speech **31**, 57–75.

Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.

Howes, D. (**1957**). "On the relationship between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–307.

Johnson, J. S., and Newport, E. L. (**1989**). "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language," Appl. Cognit. Psychol. **21**, 60–99.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 148–154.

Liberman, A. M., Cooper, F. S., Harris, K. S., and MacNeilage, P. F. (**1962**). "A motor theory of speech perception," in Proceedings of the Speech Communication Seminar, Stockholm, pp. 1–12.

Lotto, A. J., Hickok, G. S., and Holt, L. L. (**2009**). "Reflections on mirror neurons and speech perception," Trends Cogn. Sci. **13**, 110–114.

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.

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.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.

Nilsson, M., Soli, S. D., and Gelnett, D. J. (**1996**). *Development and Norming of a Hearing in Noise Test for Children* (House Ear Institute, Los Angeles, CA).

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

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., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.

Nittrouer, S., and Chapman, C. (**2009**). "The effects of bilateral electric and bimodal electric-acoustic stimulation on language development," Trends Amplif. **13**, 190–205.

Nittrouer, S., and Lowenstein, J. H. (**2009**). "Does harmonicity explain children's cue weighting of fricative-vowel syllables?," J. Acoust. Soc. Am. **125**, 1679–1692.

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. Hum. Percept. Perform. **35**, 1245–1253.

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.

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.

Revoile, S. G., Holden-Pitt, L., and Pickett, J. M. (**1985**). "Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or with normal hearing," J. Acoust. Soc. Am. **77**, 1263–1265.

Rosenzweig, M. R., and Postman, L. (**1957**). "Intelligibility as a function of frequency of usage," J. Exp. Psychol. **54**, 412–422.

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

Stelmachowicz, P. G., Lewis, D. E., Kelly, W. J., and Jesteadt, W. (**1990**). "Speech perception in low-pass filtered noise for normal and hearing-impaired listeners," J. Speech Hear. Res. **33**, 290–297.

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.
Sussman, J. E. (**1993**). "Auditory processing in children's speech perception: Results of selective adaptation and discrimination tasks," J. Speech Hear. Res. **36**, 380–395.
Uchanski, R. M., Davidson, L. S., Quadrizius, S., Reeder, R., Cadieux, J., Kettel, J., and Chole, R. A. (**2009**). "Two ears and two (or more?) devices: A pediatric case study of bilateral profound hearing loss," Trends Amplif. **13**, 107–123.
Van Engen, K. J., and Bradlow, A. R. (**2007**). "Sentence recognition in native- and foreign-language multi-talker background noise," J. Acoust. Soc. Am. **121**, 519–526.
Watson, J. M. M. (**1997**). "Sibilant-vowel coarticulation in the perception of speech by children with phonological disorder," Ph.D. thesis, Queen Margaret College, Edinburgh.
Zampini, M. L., Clarke, C., and Green, K. P. (**2000**). "Language experience and the perception of stop consonant voicing in Spanish: The case of late English-Spanish bilinguals," in *Spanish Applied Linguistics at the Turn of the Millennium*, edited by R. P. Leow and C. Sanz, (Cascadilla, Somerville, MA), pp. 194–209.

Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration

Mirjam Broersma[a)]
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands

(Received 11 September 2008; revised 13 October 2009; accepted 22 December 2009)

Does experience with a perceptual cue for a phoneme contrast in the native language affect its use in a second language for a similar contrast in a different phonetic context? Two experiments investigated Dutch and English listeners' use of preceding vowel duration as a perceptual cue for nonword-final fricative voicing in English. Dutch listeners have native language experience with the use of vowel duration for vowel length and intervocalic obstruent voicing contrasts, but not for final voicing contrasts, as Dutch does not have voiced obstruents word-finally. Previous research [Broersma, M. (2005). J. Acoust. Soc. Am. **117**, 3890–3901; (2008) **124**, 712–715] showed that Dutch listeners used vowel duration less for final /v-f/ categorization than English listeners did when vowel duration varied only between subjects, discouraging its use as a perceptual cue. The present study assessed the use of vowel duration for final /v-f/ and /z-s/ contrasts when it varied within subjects. A goodness rating and a phonetic categorization experiment showed that Dutch listeners used vowel duration, but less than English listeners did. Thus, experience with a perceptual cue for a different contrast and for a similar contrast in a different position in the native language did not lead to native-like use of this cue in the second language.
 [DOI: 10.1121/1.3292996]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.Sy [ADP] Pages: 1636–1644

I. INTRODUCTION

A lot is known about the perception of the sounds of a second language. A large body of research (see, e.g., the papers in Strange, 1995 and Bohn and Munro, 2007) has shown that mismatches between the phoneme repertoires of the first (L1) and second (L2) languages cause perceptual difficulties and that, therefore, distinguishing some L2 speech sounds is extremely difficult (Best and Strange, 1992), whereas distinguishing others is easy (Best *et al.*, 1988). Such research findings have inspired models of L2 or foreign language speech sound perception like the speech learning model (Flege, 1995) and the perceptual assimilation model (PAM) (Best *et al.*, 1988; Best and Tyler, 2007).

Not much is known, on the other hand, about the use of perceptual cues in L2 sound perception. It is not clear whether experience with the use of a perceptual cue in the L1 affects its use in the L2. In other words, can listeners "transfer" the use of an L1 perceptual cue to perception in the L2?

Experience with perceptual cues from the L1 might affect perception of similar L2 contrasts (e.g., French listeners' experience with preceding vowel duration as a cue for the final /z-s/ contrast may have affected their use of this cue for final /z-s/ decisions in English too; Flege and Hillenbrand, 1986) but it might also aid perception of completely different contrasts. Crowther and Mann (1992) found that native listeners of Japanese, which distinguishes between long and short vowels, were also sensitive to vowel duration when they categorized English final /d/ and /t/. Their experience with vowel duration as a perceptual cue for Japanese vowel contrasts might thus have aided their perception of English voiced and voiceless final stops.

Experience with perceptual cues from the L1 may, however, not always affect perception of L2 contrasts, even when the perceptual cue is used for similar L1 and L2 contrasts but in a different phonetic position. Broersma (2005, 2008) found that native listeners of Dutch, which has /v-f/ and /z-s/ contrasts similar to the English contrasts but does not distinguish between voiced and voiceless obstruents in word-final position, used the duration of the preceding vowel less consistently than English listeners did. Dutch listeners have native language experience with the use of vowel duration as a perceptual cue for vowel length contrasts and for intervocalic consonant voicing; yet, they did not seem to apply this experience when categorizing English final fricatives.

This paper further investigates the transfer of the use of perceptual cues from the L1 to the L2. It investigates to which extent Dutch listeners, who are familiar with the use of vowel duration as a perceptual cue, but not for final voicing contrasts (as their L1 does not have those), use vowel duration for this purpose when listening to English. To that end, the study of Broersma (2005, 2008) was extended with a different experimental design and a different paradigm.

In English, vowels are generally longer before a voiced final fricative than before a voiceless final fricative (Jones, 1950), and the duration of the preceding vowel affects listeners' perception of final fricatives as voiced or voiceless (Denes, 1955). Duration of the preceding vowel is a robust and very important perceptual cue for final fricative voicing for English listeners (Derr and Massaro, 1980; Hogan and Rozsypal, 1980; Raphael, 1972).

[a)]Electronic mail: mirjam@mirjambroersma.nl

Like English, Dutch has voiced and voiceless obstruents, namely, /b-p/, /d-t/, /v-f/, /z-s/, and /k/. All of the voiceless and none of the voiced obstruents can occur at the end of words in isolation (Booij, 1995). Dutch thus has a /z-s/ and a /v-f/ contrast, but does not have those contrasts in word-final position. Dutch listeners nevertheless do have some native language experience with the use of vowel duration for fricative voicing. For intervocalic two-obstruent sequences in Dutch (with the first, the second, or both being a fricative), duration of the preceding vowel plays a small but significant role for voicing perception (Van den Berg, 1987, 1988, 1989). For single intervocalic fricative voicing, no effect of vowel duration has been demonstrated on perception (e.g., Slis and Cohen, 1969b; Slis and Van Heugten, 1989), but vowels preceding a voiced intervocalic fricative are on average 40 ms longer than those preceding a voiceless intervocalic fricative in Dutch (Slis and Cohen, 1969a). Note that this difference for intervocalic fricatives in Dutch is much smaller than that before final fricatives in English; e.g., Peterson and Lehiste (1960) showed an average difference of 148 ms before voiced and voiceless final fricatives. Finally, results by Jongman *et al.* (1992) suggest that Dutch listeners can generalize their knowledge about the relation between vowel duration and intervocalic obstruent voicing to the case of word-final voicing, as the underlying voicing of voiceless word-final obstruents affects the way Dutch listeners categorize the duration of preceding vowels. Further, Dutch also has phonemically long and short vowels that are mainly distinguished by phonetic vowel duration (Booij, 1995). Thus, Dutch listeners have native language experience with the use of vowel duration as a perceptual cue both for obstruent voicing and vowel length contrasts.

The Dutch listeners studied by Broersma (2005, 2008) had a high level of proficiency in English as a second language and ample exposure to English. They had started receiving English instruction in primary school and were regularly exposed to English, i.e., through the media and in the educational system, as is common in The Netherlands. They therefore seem well equipped for a native-like use of preceding vowel duration as a perceptual cue for the English final /v-f/ and /z-s/ contrasts. Yet, the results of Broersma (2005, 2008) showed robust differences between the Dutch and the English listeners' use of vowel duration as a final fricative voicing cue in English.

In the experiment of Broersma (2005, 2008), vowel duration was made uninformative, and whereas the Dutch listeners initially used vowel duration to the same extent as the English listeners did, they soon stopped using it, whereas the English listeners kept using it persistently throughout the experiment. Thus, the Dutch but not the English listeners adapted their use of vowel duration to the nature of the experimental materials. Vowel duration was made uninformative to preclude the possibility that the Dutch listeners would use vowel duration only as a result of the stimulus materials, as in experiments with a limited number of variables, listeners may rely more on those variables than they would do with normal speech (Bohn, 1995; Bradlow, 1996). Thus, each participant heard either a phonetically long or a phonetically short vowel throughout the experiment, combined with a fricative from an 11-step voiced to voiceless continuum. Because for each participant vowel duration was kept constant, it was not an informative cue for final fricative voicing. Moreover, vowel duration mismatched the voicing information in the fricatives for some of the items (i.e., for /v/-like fricatives preceded by a short vowel and for /f/-like fricatives preceded by a long vowel).

Broersma (2005) showed that Dutch listeners used vowel duration less than English listeners did as a perceptual cue for the final /v-f/ contrast. Broersma (2008) further showed that the difference between the Dutch and English listeners' results was not immediately present, but arose during the first part of the experiment. In the 44 trials of the practice part of the experiment, the Dutch and the English listeners used vowel duration to a similar extent. After that, vowel duration did not affect Dutch listeners' responses anymore, whereas its effect remained unchanged for the English listeners' responses. For the English listeners, vowel duration strongly affected the responses, such that listeners in the long vowel condition gave more voiced responses than listeners in the short vowel condition did. Long vowels affected their categorization responses especially strongly; long vowels combined with fricatives at the voiceless end point of the continuum received 31% "V" responses. It was concluded that the Dutch listeners adapted to the nature of the stimulus materials more flexibly than the English listeners, who had extensive native language experience with vowel duration as a cue for final fricative voicing. The difference between the Dutch and the English listeners was thus at least partly related to the fact that vowel duration was kept constant for each participant due to the between-subjects design of the experiment.

Another indication that the experimental design induced a limited use of vowel duration as a perceptual cue for voicing are the categorization results for the final /z-s/ contrast (Broersma, 2005). For this contrast, neither the English nor the Dutch listeners' responses showed an effect of vowel duration, while previous research has shown that vowel duration is an important cue for final voicing distinctions for this contrast at least for English listeners (e.g., Derr and Massaro, 1980; Hogan and Rozsypal, 1980; Raphael, 1972). The question thus remains to which extent the results were due to the experimental design, and if there are also differences between Dutch and English listeners' use of vowel duration for the final /v-f/ contrast, and possibly for the final /z-s/ contrast, if vowel duration is not kept constant for each participant.

The present paper investigates Dutch and English listeners' use of vowel duration for final fricative voicing in English, with an experimental design that does not discourage the use of vowel duration as a perceptual cue. The same stimuli are used as in Broersma (2005, 2008), but now, instead of blocking vowel duration per participant, all participants are presented with both vowel durations (and with the 11 steps of the fricative voicing continuum, as in the earlier experiment). In Experiment 1, Dutch and English listeners' goodness ratings of the stimuli are collected. In Experiment 2, Dutch and English listeners' phonetic categorization of the

same stimuli is assessed. Experiment 2 is thus a replication of Broersma (2005, 2008), except that now, crucially, vowel duration is varied within subjects.

Both experiments address three questions. First, is there a difference between the Dutch and the English listeners' use of vowel duration for the /v-f/ contrast when it is an informative perceptual cue within the experiment? Second, is vowel duration used as a perceptual cue for the /z-s/ contrast as well, unlike in Broersma (2005, 2008) and, third, if so, do Dutch and English listeners use vowel duration differentially for the /z-s/ contrast?

II. EXPERIMENT 1

A. Method

1. Participants

Participants were 16 native speakers of Dutch and 16 native speakers of British English. None of them had participated in the experiments reported in Broersma (2005), but they were part of the same populations tested there. The Dutch participants were proficient in English as a second language. They had received on average 7 years of English instruction in primary and secondary schools and were regularly exposed to English through the media and at the university they attended. The English participants did not know any Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, and the English participants from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex. None reported any speech or hearing disorders. All were volunteers and received a small fee for participation.

2. Materials

The same materials were used as in Broersma (2005). A male native speaker of British English recorded two tokens of the nonwords /ku:v/, /ku:f/, /fu:z/, and /fu:s/, in a clear citation style, in a soundproof booth. The materials were recorded with a Sennheiser microphone onto digital audiotape and downsampled to 16 kHz during transfer to a computer. Further editing was done with XWAVES software.

For each nonword, from one token, the final fricative was extracted. Thus, /v/, /f/, /z/, and /s/ were extracted from the end of each nonword, with the cut being made at the last positive zero crossing before the onset of frication noise. Two continua, one from /v/ to /f/ and one from /z/ to /s/, were created following the procedure of Stevenson (1979) and Repp (1981). The continua consisted of 11 steps, ranging from the natural voiced end point /v/ or /z/ (step 1) to the natural voiceless end point /f/ or /s/ (step 11), with 9 intermediate steps. To this end, first, the final /f/ and /s/ were shortened to match the durations of the final /v/ and /z/.[1] They were shortened by 56–127 ms (31%) for the /f/ and by 80–187 ms (30%) for the /s/ by removing a portion from the center of the fricative. Next, for the pairs of phonemes thus obtained, the amplitudes of the waveforms were added in varying proportions in 11 equally spaced steps (Stevenson, 1979; Repp, 1981).

For each nonword, a carrier containing all but the fricative was extracted from the other token. Thus, from the nonwords /ku:v/ and /ku:f/, /ku:/ was extracted, and from the nonwords /fu:z/ and /fu:s/, /fu:/ was extracted, removing the final fricative, truncating the signal at the last positive zero crossing before the start of the frication noise. For each continuum, there were thus two carriers. One carrier was originally pronounced with a voiced final fricative and contained a phonetically long vowel (of 257 ms for the /v-f/ and 233 ms for the /z-s/ contrast) and the other was originally pronounced with a voiceless final fricative and contained a phonetically short vowel (of 98 ms for the /v-f/ and 118 ms for the /z-s/ contrast).

For each continuum, the two end points and the nine intermediate steps were spliced onto the end of the appropriate carriers. Thus, for the /v-f/ continuum, all steps were spliced onto both /ku:/ carriers, and for the /z-s/ continuum, all steps were spliced onto both /fu:/ carriers. Auditory examination ensured that the manipulation had not resulted in audible discontinuities.

3. Design

Stimuli were blocked by contrast but not by vowel duration. There were four blocks, two for the /v-f/ contrast and two for the /z-s/ contrast. For the /v-f/ items, participants decided in one block to which extent each fricative sounded like a good "V," and in the other block to which extent the same items sounded like a good "F," and for the /z-s/ items, participants decided in one block to which extent each fricative sounded like a good "Z," and in the other block like a good "S." As the carrier for the /z-s/ contrast contained an /f/, the /v-f/ contrast was always tested before the /z-s/ contrast; otherwise the order of the blocks was counterbalanced. Each block contained all combinations of 11 fricatives and two carriers presented in random order.

4. Procedure

Participants were tested one at a time in a quiet room. They received written instructions in their native language that they would hear a series of English nonwords, and they were asked to pay attention to the final sound. Before each block, they were informed about the target sound in that block. Participants were instructed to decide whether the final sound of each stimulus was a very poor, fairly poor, fairly good, or very good instance of the target sound and to indicate their response by pressing one of four buttons, labeled from "1—poor" to "4—good."

The experiment started with a short practice during which the participants heard nonwords ending with a /b/ or a /p/, some of them unedited and others with the release burst removed (Broersma 2005, Experiment 1) and indicated to what degree this sound corresponded to a "B."

The experiment was controlled with Nijmegen Experiment Set-Up (NESU) software, and stimuli were presented binaurally, one at a time, over Sennheiser closed headphones at a comfortable listening level. Participants responded by pressing one of four response buttons. Stimuli were presented with 3000 ms intervals between onsets.

B. Results and discussion

Analyses of variance (ANOVAs) were done on averaged goodness ratings, with vowel duration and fricative voicing as within-subjects factors and native language as between-subjects factor. As a measure of effect size, *partial* η^2 is reported. A *partial* η^2 of 0.01 is considered to indicate a small effect size, a *partial* η^2 of 0.06 a medium, and of 0.14 a large effect size. Mean responses are presented in Table I.

First, for the "V" target, Fig. 1(a) shows that both for the Dutch and for the English listeners, as expected, goodness ratings are higher for items with a long vowel than for items with a short vowel [vowel duration: $F(1,30)=209.23, p<0.001$, *partial* $\eta^2=0.88$] and higher for items with more voiced fricatives than for items with less voiced fricatives [fricative voicing: $F(10,300)=8.86, p<0.001$, *partial* $\eta^2=0.23$]. Crucially, however, the use of vowel duration differs for the Dutch and English listeners in part of the voicing continuum, as reflected in a significant interaction among vowel duration, fricative voicing, and native language [$F(10,300)=2.79, p<0.01$, *partial* $\eta^2=0.09$]. For the English listeners, in the condition with the short preceding vowel, goodness ratings are higher for items with more voiced fricatives than for items with less voiced fricatives, as expected [$F(10,150)=9.36, p<0.001$, *partial* $\eta^2=0.38$], but in the condition with the long preceding vowel, there is no significant effect of fricative voicing [$F(10,150)=1.49, p>0.1$]. Thus, English listeners give stimuli with more voiced fricatives and stimuli with less voiced fricatives similar goodness ratings for "V" when they are preceded by a long vowel. For those listeners, indeed, there is an interaction between vowel duration and fricative voicing [$F(10,150)=4.97, p<0.001$, *partial* $\eta^2=0.25$]. Thus, long vowel durations affect the English listeners' goodness ratings more than short vowel durations do. For the Dutch listeners, on the other hand, there is no interaction between vowel duration and fricative voicing [$F(10,150)<1$]. For those listeners, goodness ratings are higher for items with more voiced fricatives than for items with less voiced fricatives [$F(10,150)=2.88, p<0.01$, *partial* $\eta^2=0.16$] for both vowel durations alike, and goodness ratings are higher for items with a long vowel than for items with a short vowel [$F(1,15)=91.70, p<0.001$, *partial* $\eta^2=0.86$]. This pattern is in line with the results reported in Broersma (2005, 2008) that for English listeners, when vowel duration and fricative voicing do not match, the long vowel but not the short vowel outweighs fricative voicing, but for Dutch listeners, long and short vowels affect /v-f/ categorization to the same extent.

Second, for the "F" target, Fig. 1(b) shows that for Dutch and English listeners, as expected, goodness ratings are higher for items with a short vowel than for items with a long vowel [$F(1,30)=27.13, p<0.001$, *partial* $\eta^2=0.48$] and higher for items with less voiced fricatives than for items with more voiced fricatives [$F(10,300)=14.69, p<0.001$, *partial* $\eta^2=0.33$]. Crucially, vowel duration affects Dutch listeners' goodness ratings for "F" less than the English listeners' ratings, as reflected in an interaction between vowel duration and native language [$F(1,30)=4.50, p<0.05$, *partial* $\eta^2=0.13$]. For the English listeners, the effect of vowel duration, in the expected direction, is significant [$F(1,15)=103.03, p<0.001$, *partial* $\eta^2=0.87$]; for the Dutch listeners, it is not [$F(1,15)=2.74, p>0.1$]. For the condition with the short preceding vowel, there is no difference between the Dutch and English' listeners' goodness ratings, [$F(1,30)<1$], but for the condition with the long preceding vowel, Dutch listeners give higher scores than English listeners do [$F(1,30)=11.01, p<0.01$, *partial* $\eta^2=0.27$]. Thus, not all vowel durations but specifically long vowel durations lead to different ratings for English listeners than for Dutch listeners. This is in line with the finding for the "V" target that long vowels affect the English listeners but not the Dutch listeners particularly strongly.

Comparing "V" and "F" targets, in the condition with the long vowel, Dutch listeners give high goodness ratings for the "V" target, but also relatively high goodness ratings for the "F" target. Thus, whereas the English listeners perceive stimuli with a long vowel as a good "V" and as a poor "F," the Dutch listeners perceive them as good exemplars of "V" but also as relatively good exemplars of "F", showing again that vowel duration is not as decisive for the Dutch listeners' perception of the goodness of the voicing of those fricatives as it is for the English listeners' perception.

For the "Z" target [Fig. 2(a)] and for the "S" target [Fig. 2(b)], crucially, vowel duration significantly affects goodness ratings too. For the "Z" target, items with a long vowel receive higher goodness ratings than items with a short vowel [$F(1,30)=68.43, p<0.001$, *partial* $\eta^2=0.70$], and for the "S" target, items with a short vowel receive higher goodness ratings than items with a long vowel [$F(1,30)=50.40, p<0.001$, *partial* $\eta^2=0.63$] for Dutch and English listeners alike. Further, the effect of fricative voicing is also as expected: for the "Z" target, goodness ratings are higher for items with more voiced fricatives than for items with less voiced fricatives [$F(10,300)=31.03, p<0.001$, *partial* $\eta^2=0.51$], and for the "S" target, goodness ratings are higher for items with less voiced fricatives than for items with more voiced fricatives [$F(10,300)=32.36, p<0.001$, *partial* $\eta^2=0.52$] for Dutch and English listeners alike.

Finally, comparing the /v-f/ and the /z-s/ contrast, there is an interaction between vowel duration and contrast [$F(1,30)=4.39, p<0.05$, *partial* $\eta^2=0.13$], confirming that the effect of vowel duration is larger for the /v-f/ contrast than for the /z-s/ contrast.

In summary, for the /v-f/ and the /z-s/ contrast, both vowel duration and fricative voicing affect the goodness ratings as expected: long vowels lead to higher goodness ratings than short vowels for voiced targets and vice versa for voiceless targets, and more voiced fricatives lead to higher goodness ratings than more voiceless fricatives for voiced targets and vice versa for voiceless targets. Effect sizes are large both for vowel duration and fricative voicing, but larger for the former than for the latter. Importantly, the effect of vowel duration is found not only for the /v-f/ contrast but also for the /z-s/ contrast.

For the /v-f/ contrast, vowel duration and, in particular, long vowels affect English listeners' goodness ratings more than Dutch listeners' ratings. These results are in line with the phonetic categorization results from Broersma (2005,

TABLE I. Results Experiments 1 and 2.

	Step fricative voicing continuum	Mean goodness ratings			
		Dutch participants		English participants	
		Long vowel	Short vowel	Long vowel	Short vowel
Experiment 1					
"V" target	1(voiced)	3.8	2.2	3.8	2.4
	2	3.6	1.9	3.6	2.8
	3	3.7	2.4	3.6	2.7
	4	3.3	2.0	3.7	2.2
	5	3.6	1.8	3.4	1.9
	6	3.3	1.8	3.7	1.6
	7	3.4	1.9	3.5	1.6
	8	3.0	1.7	3.4	1.6
	9	3.1	1.8	3.5	1.3
	10	3.4	1.8	3.3	1.3
	11	3.1	1.9	3.3	1.4
"F" target	1(voiced)	2.1	2.3	1.3	2.0
	2	2.1	2.3	1.3	2.2
	3	2.3	2.7	1.4	2.4
	4	2.1	2.6	1.6	2.6
	5	2.1	2.8	1.5	2.8
	6	2.4	2.9	1.5	3.1
	7	2.1	3.1	1.6	3.0
	8	2.9	3.4	1.6	3.4
	9	2.4	3.3	1.4	3.4
	10	2.7	3.3	1.9	3.3
	11	2.6	3.2	1.9	3.1
"Z" target	1(voiced)	3.9	3.4	3.6	2.7
	2	3.6	2.4	3.4	2.6
	3	3.4	2.6	3.4	2.5
	4	3.4	2.6	3.3	2.1
	5	2.9	2.6	3.2	2.2
	6	3.3	1.8	3.0	1.5
	7	3.2	1.8	2.8	1.6
	8	2.4	1.3	2.8	1.4
	9	2.3	1.4	2.3	1.2
	10	2.1	1.3	2.4	1.3
	11	2.2	1.4	2.4	1.3
"S" target	1(voiced)	1.3	1.8	2.0	2.1
	2	1.3	2.0	1.4	2.6
	3	1.5	2.1	1.7	2.2
	4	1.4	2.1	1.8	2.4
	5	1.9	2.6	2.1	2.7
	6	1.9	2.8	2.1	2.9
	7	2.1	3.3	2.3	3.4
	8	1.9	3.3	2.3	3.4
	9	2.3	3.4	2.6	3.8
	10	2.6	3.9	2.9	3.7
	11	2.6	3.9	2.4	3.6

	Step fricative voicing continuum	Mean percentage "voiced" responses			
		Dutch participants		English participants	
		Long vowel	Short vowel	Long vowel	Short vowel
Experiment 2					
/v-f/ contrast	1(voiced)	93.4	36.9	91.6	32.8
	2	94.7	34.1	93.8	30.6
	3	92.2	31.9	95.9	25.0
	4	93.1	23.1	95.6	18.1
	5	90.6	15.6	95.0	11.9
	6	88.1	12.5	94.4	6.3
	7	83.1	8.4	88.1	8.4
	8	77.2	5.3	87.8	6.9
	9	69.1	6.3	87.2	8.4
	10	67.8	4.7	88.8	6.6
	11	69.1	5.0	85.3	5.6
/z-s/ contrast	1(voiced)	92.5	58.4	92.5	38.4
	2	95.0	55.3	90.0	31.3
	3	94.4	52.8	88.8	31.6
	4	90.9	51.9	88.1	29.7
	5	88.4	37.2	86.6	24.4
	6	84.4	22.8	79.7	19.7
	7	74.1	13.4	78.4	10.6
	8	65.3	8.1	66.6	6.6
	9	61.6	5.9	65.0	5.9
	10	57.2	5.3	62.2	5.0
	11	56.6	2.5	62.5	5.3

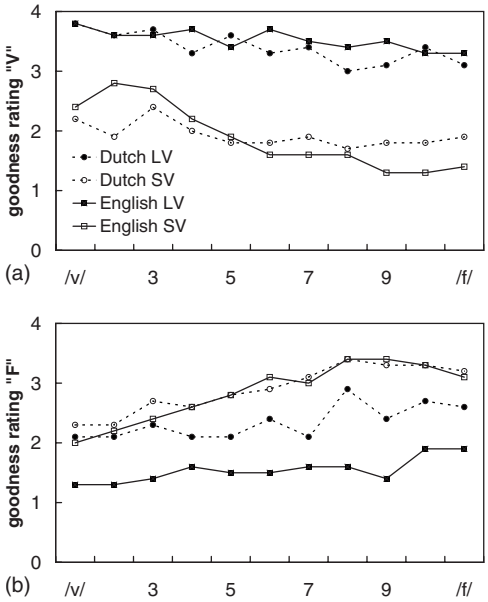


FIG. 1. Mean goodness rating as a function of the place on an 11-step /v/ to /f/ continuum, preceding vowel duration (LV: long vowel; SV: short vowel), and participants' native language: (a) target "V" and (b) target "F."

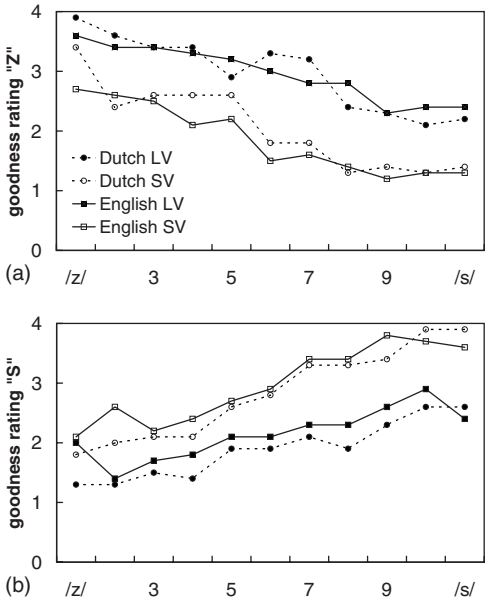


FIG. 2. Mean goodness rating as a function of the place on an 11-step /z/ to /s/ continuum, preceding vowel duration (LV: long vowel; SV: short vowel), and participants' native language: (a) target "Z" and (b) target "S."

2008) that showed a larger effect of vowel duration for English listeners than for Dutch listeners and an especially large effect of long vowel durations when combined with voiceless fricatives for English listeners. The present results show that those findings were not an artifact of the design of Broersma (2005, 2008), where vowel duration was kept constant for each participant, but is also found when vowel duration is varied for each participant.

Experiment 2 further investigates the use of vowel duration when it varies within subjects with a phonetic categorization task. The experiment is identical to that in Broersma (2005, 2008), except that all participants hear both the phonetically long vowel and the phonetically short vowel. Stimuli thus vary in vowel duration and in fricative voicing for each participant. Cue trading between temporal and spectral cues has often been found (Fitch *et al.*, 1980; Sinnott and Saporita, 2000), and a trading relation between vowel duration and fricative voicing is anticipated here too. Note that whereas fricative voicing varies in 11 steps, vowel duration varies in only two steps, which might make the latter variable relatively salient (Bohn, 1995). The experiment addresses the same three questions; first, whether there is still a difference between the Dutch and the English listeners' use of vowel duration for the /v-f/ contrast when it varies within subjects, second, whether vowel duration is used as a perceptual cue for the /z-s/ contrast, like in Experiment 1, but unlike in Broersma (2005, 2008), and third, whether Dutch and English listeners use vowel duration differentially for the /z-s/ contrast.

III. EXPERIMENT 2

A. Method

1. Participants

Participants were 16 native speakers of Dutch and 16 native speakers of British English, matching the description given for Experiment 1, none of whom had participated in Experiment 1.

2. Materials

The same materials were used as in Broersma (2005) and in Experiment 1.

3. Design

As in Experiment 1, stimuli were blocked by contrast but, unlike in Broersma (2005), not by vowel duration. Thus, each participant heard both the phonetically long and the phonetically short vowel within the same experiment and within the same block. As the carrier for the /z-s/ contrast contained an /f/, the /v-f/ contrast was always tested before the /z-s/ contrast. Each block contained 20 repetitions of all combinations of 11 fricatives and two carriers, yielding 440 trials per block. Items were semi-randomized, such that the same step of the fricative continuum could not occur twice in succession. There was no practice with the crucial contrasts, only with the /b-p/ contrast (as in Experiment 1).

4. Procedure

Participants were tested one at a time in a quiet room. They received written instructions in their native language that they would hear a series of English nonwords, which would be similar except for the final sound. They were instructed to decide whether this was a "V" or an "F," or a "Z" or an "S," respectively, and to indicate their response by pressing one of two buttons, labeled "V" and "F," or "Z" and "S." The experiment started with a short practice during which the participants categorized final stops as "B" or "P."

The experiment was controlled with NESU experimental software. Stimuli were presented binaurally, one at a time, over Sennheiser closed headphones at a comfortable listening level. Participants responded by pressing one of two response buttons. No time limit was imposed for the responses. After each button press, presentation of the next item started.

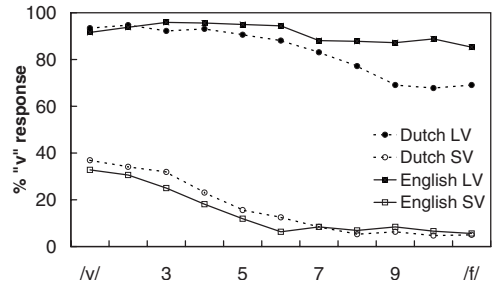


FIG. 3. Mean percentage of "V" responses as a function of the place on an 11-step /v/ to /f/ continuum, preceding vowel duration (LV: long vowel; SV: short vowel), and participants' native language.

B. Results and discussion

Two responses with reaction times longer than 10 000 ms were removed. As Figs. 3 and 4 show, for both contrasts, the effect of vowel duration is very large. The categorization curves are very shallow, and the mean percentages of "voiced" responses for the conditions with the long and short preceding vowels do not overlap, except for the Dutch listeners' responses for the /z-s/ contrast, where they just touch (with 57% "Z" responses for step 11 in the condition with the long preceding vowel and 58% for step 1 in the condition with the short preceding vowel). ANOVAs were done on arcsine transformed proportions of "voiced" responses, with vowel duration and fricative voicing as within-subjects factors and native language as between-subjects factor. Mean responses are presented in Table I.

Figure 3 shows that especially for the /v-f/ contrast, the effect of vowel duration is extremely large and the effect of fricative voicing is relatively small, both for the Dutch and for the English listeners. Indeed, stimuli with long vowels receive more "V" responses than stimuli with short vowels, both for the Dutch listeners [$F(1,15)=59.43, p<0.001$, *partial* $\eta^2=0.80$] and for the English listeners [$F(1,15)=160.95, p<0.001$, *partial* $\eta^2=0.92$]. Items with more voiced fricatives also receive significantly more "V" responses than items with less voiced fricatives for the Dutch listeners [$F(10,150)=12.83, p<0.001$, *partial* $\eta^2=0.46$] and for the English listeners [$F(10,150)=6.76, p<0.001$, *partial* $\eta^2=0.31$].

Crucially, however, there are differences between the Dutch and English listeners' categorization responses, reflected in a significant three-way interaction among vowel duration, fricative voicing, and native language [$F(10,300)=1.90, p<0.05$, *partial* $\eta^2=0.06$]. As Fig. 3 shows, in the condition with the long preceding vowel, the English listeners give more "V" responses than the Dutch listeners to items with less voiced fricatives. Indeed, in the condition with the long preceding vowel, the interaction between fricative voicing and native language is significant [$F(10,300)=2.60, p<0.005$, *partial* $\eta^2=0.08$]. For the condition with the short preceding vowel, there is no interaction between fricative voicing and native language [$F(10,300)<1$].

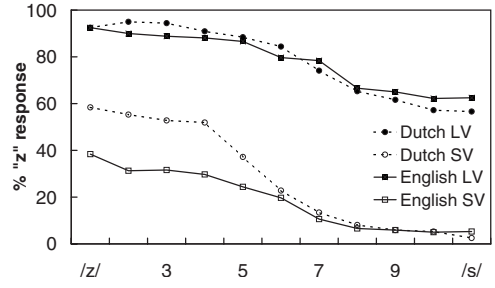


FIG. 4. Mean percentage of "Z" responses as a function of the place on an 11-step /z/ to /s/ continuum, preceding vowel duration (LV: long vowel; SV: short vowel), and participants' native language.

For the /z-s/ contrast, Fig. 4 again shows a large effect of vowel duration and a relatively small effect of fricative voicing, both for the Dutch and for the English listeners. Similar to the /v-f/ contrast, stimuli with long vowels receive more "Z" responses than stimuli with short vowels, both for the Dutch listeners [$F(1,15)=33.87, p<0.001$, *partial* $\eta^2=0.69$] and for the English listeners [$F(1,15)=61.49, p<0.001$, *partial* $\eta^2=0.80$]. Items with more voiced fricatives also receive significantly more "Z" responses than items with less voiced fricatives for the Dutch listeners [$F(10,150)=20.58, p<0.001$, *partial* $\eta^2=0.58$] and for the English listeners [$F(10,150)=14.21, p<0.001$, *partial* $\eta^2=0.49$].

Also similar to the /v-f/ contrast, there is a difference between the Dutch and English listeners' categorization responses, reflected in a significant three-way interaction among vowel duration, fricative voicing, and native language [$F(10,300)=2.09, p<0.05$, *partial* $\eta^2=0.07$]. Figure 4 shows that the difference between the Dutch and English listeners' responses is now located on the other side of the graph than for the /v-f/ contrast: in the condition with the short preceding vowel the English listeners give fewer "Z" responses than the Dutch listeners to items with more voiced fricatives. In that condition, the interaction between fricative voicing and native language is significant [$F(10,300)=2.98, p<0.001$, *partial* $\eta^2=0.09$]. For the condition with the long preceding vowel, there is no interaction between fricative voicing and native language [$F(10,300)<1$].

The results for the /v-f/ and /z-s/ contrasts are thus to a large extent similar: both Dutch and English listeners use vowel duration and fricative voicing as a perceptual cue, effect sizes are large both for vowel duration and fricative voicing but larger for the former than for the latter, and for both contrasts, differences between the Dutch and English listeners' responses appear where the stimuli contain mismatching cues. However, where these differences occur is different for both contrasts. For the /v-f/ contrast, English listeners give more "voiced" responses than the Dutch listeners do to items with less voiced fricatives preceded by a long vowel, and for the /z-s/ contrast, English listeners give more "voiceless" responses than the Dutch listeners do to items with more voiced fricatives preceded by a short vowel. Thus, in these places, the English listeners rely more on vowel duration than the Dutch listeners do.[2]

The results for the /v-f/ contrast are in line with those from Broersma (2005), where English listeners also gave more "voiced" responses than Dutch listeners did to long vowels combined with less voiced fricatives. For the /z-s/ contrast, on the other hand, there was no effect of vowel duration in that study. In the present study, the effect of

vowel duration is also smaller and the effect of fricative voicing is larger for the /z-s/ contrast than for the /v-f/ contrast (an ANOVA with contrast as an additional within-subjects factor shows significant interactions between vowel duration and contrast [$F(1,30)=11.37, p<0.01$, *partial* $\eta^2=0.28$)] and between fricative voicing and contrast [$F(10,300)=5.90, p<0.001$, *partial* $\eta^2=0.16$]). However, the effect size of vowel duration for the /z-s/ contrast is still very large, and crucially, vowel duration affects Dutch and English listeners differentially.

IV. GENERAL DISCUSSION

Three questions were raised in the Introduction. First, would Dutch and English listeners use vowel duration differently for the /v-f/ contrast when it was varied for each participant? Both experiments showed that this was indeed the case. Thus, the finding of Broersma (2005, 2008) that Dutch listeners used vowel duration less than English listeners did was not an artifact of the between-subjects manipulation of vowel duration in that study. In the present experiments, where the design did not discourage the use of vowel duration as a perceptual cue, Dutch listeners still used it less than English listeners did.

The second question was whether there would be an effect of vowel duration for the /z-s/ contrast when vowel duration varied within subjects. Indeed, in both experiments, a clear effect of vowel duration was found. Thus, the lack of such an effect in Broersma (2005) was most likely due to the design of that study. The finding that the English listeners used vowel duration for the final /z-s/ distinction is in line with earlier research (e.g., Derr and Massaro, 1980; Hogan and Rozsypal, 1980; Raphael, 1972). The effect of vowel duration was smaller for the /z-s/ contrast than for the /v-f/ contrast in both experiments.

The third question was whether Dutch and English listeners would use vowel duration differently for the /z-s/ contrast too. In Experiment 1, there were no differences between the Dutch and English listeners' use of vowel duration for the /z-s/ contrast. Experiment 2 (involving a more sensitive paradigm) showed, however, that English listeners used vowel duration more than Dutch listeners did. Thus, the differences between Dutch and English listeners' use of vowel duration as a perceptual cue are not limited to the /v-f/ contrast, but occur for the /z-s/ contrast too.

Experiment 2 was a replication of Broersma (2005, 2008), except for the within-subjects manipulation of vowel duration. The results show that this design choice affected the listeners' use of vowel duration as expected. In Experiment 2, a large effect size of vowel duration was found for both contrasts. (This was also the case in Experiment 1, where vowel duration was also varied for each participant.) In Broersma (2005, 2008), on the other hand, where vowel duration was varied only between subjects, the effect size of vowel duration for the /v-f/ contrast was medium (with a *partial* η^2 of 0.12), and there was no significant effect of vowel duration for the /z-s/ contrast.

What do these results mean in terms of perceptual accuracy? The differences between Dutch and English listeners, although very robust, were also especially prominent where perceptual cues mismatched. Further, Dutch listeners did use vowel duration as a final voicing cue, and the differences between the Dutch and the English listeners' use of vowel duration were relatively small compared to the overall size of the effect of vowel duration. This may explain why in a previous study (Broersma, 2005, Experiment 1) Dutch listeners were found to categorize unedited tokens of final voiced and voiceless fricatives as accurately as English listeners did and as accurately as word-initial fricatives. Their use of vowel duration might be sufficiently adequate to recognize final voiced and voiceless fricatives under normal circumstances (i.e., when perceptual cues do not mismatch). Further, they might successfully use other temporal, spectral, and power properties of the stimuli (e.g., Jongman *et al.*, 2000; Silbert and De Jong, 2008) to distinguish between voiced and voiceless final fricatives. Therefore, in normal language use, the differences in cue weighting demonstrated here might not hinder Dutch listeners' comprehension of English speech much, in line with the predictions of the PAM (Best *et al.*, 1988; Best and Tyler, 2007).

Finally, what do these results mean for the transfer of the use of perceptual cues from the L1 to the L2? It was argued in the Introduction that Dutch listeners were well equipped for the use of vowel duration as a perceptual cue for final fricative voicing. Dutch provides experience with the use of vowel duration as a perceptual cue for vowel length contrasts and even for intervocalic obstruent voicing contrasts. Despite their native language experience with the use of vowel duration as a perceptual cue, Dutch listeners used it less than English listeners did for the English final fricative voicing contrasts. Thus, the Dutch listeners (advanced L2 learners with ample exposure to English) did not seem to apply their L1 experience with vowel duration when listening to English contrasts. Whether their L1 experience affected the use of vowel duration for final fricative voicing at all, the present data cannot reveal, but the results showed that the Dutch listeners did not rely on this cue as strongly as native English listeners did. The findings extend those from Broersma (2005, 2008) and seem to reflect a robust difference between Dutch and English listeners' use of vowel duration that is not limited to a single phoneme contrast or to a particular experimental design or paradigm. Thus, experience with a perceptual cue for a contrast in the L1 may not suffice for the efficient use of this cue for a different contrast in the L2, or even for a similar contrast in a different phonetic position in the L2.

ACKNOWLEDGMENTS

This research was supported by a Veni grant from the Netherlands Organisation for Scientific Research (NWO). Many thanks to Alan Garnham of the Laboratory of Experimental Psychology, University of Sussex, and to Anne Cutler of the Max Planck Institute for Psycholinguistics, Nijmegen, for the opportunity to test participants in their laboratories. I am grateful to Marco van de Ven for testing Dutch participants and to an anonymous reviewer for helpful comments.

[1]Voiced and voiceless fricatives of equal duration were needed in order to create the continua. As frication noise is generally longer for voiceless final fricatives than for voiced final fricatives (Crystal and House, 1988; Silbert and De Jong, 2008) and listeners use frication duration as a perceptual cue for final fricative voicing (Watson, 1983), shortening the voiceless fricatives may have made them perceptually more voiced.

[2]In Broersma (2008), English listeners' use of vowel duration changed during the experiment. Here, this was not the case. The data for each contrast were split into seven parts. In an ANOVA with part as an additional within-subjects factor, there were no interactions with or main effect of part, showing that the use of vowel duration or fricative voicing did not change during the experiment.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.

Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 279–303.

Bohn, O.-S. and Munro, M. J., eds. (**2007**). *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (John Benjamins, Amsterdam).

Booij, G. (**1995**). *The Phonology of Dutch* (Oxford University Press, Oxford).

Bradlow, A. R. (**1996**). "A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and Spanish: Universal and language-specific aspects," Phonetica **53**, 55–85.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Broersma, M. (**2008**). "Flexible cue use in nonnative phonetic categorization," J. Acoust. Soc. Am. **124**, 712–715.

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.

Crystal, T. H., and House, A. S. (**1988**). "A note on the durations of fricatives in American English," J. Acoust. Soc. Am. **84**, 1932–1935.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Derr, M. A., and Massaro, D. W. (**1980**). "The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction," Percept. Psychophys. **27**, 51–59.

Fitch, H. L., Halwes, T., Erickson, D. M., and Liberman, A. M. (**1980**). "Perceptual equivalence of two acoustic cues for stop-consonant manner," Percept. Psychophys. **27**, 343–350.

Flege, J. E. (**1995**). "Second language learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–272.

Flege, J. E., and Hillenbrand, J. (**1986**). "Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English," J. Acoust. Soc. Am. **79**, 508–517.

Hogan, J. T., and Rozsypal, A. J. (**1980**). "Evaluation of vowel duration as a cue to the voicing distinction in the following word-final consonant," J. Acoust. Soc. Am. **67**, 1764–1771.

Jones, D. (**1950**). *The Phoneme: Its Nature and Use* (W. Heffer & Sons, Cambridge).

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (**1992**). "The phonological representation of [voice] in speech perception," Lang Speech **35**, 137–152.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Repp, B. H. (**1981**). "Perceptual equivalence of two kinds of ambiguous speech stimuli," Bull. Psychon. Soc. **18**, 12–14.

Silbert, N., and De Jong, K. (**2008**). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," J. Acoust. Soc. Am. **123**, 2769–2779.

Sinnott, J. M., and Saporita, T. A. (**2000**). "Differences in American English, Spanish, and monkey perception of the say-stay trading relation," Percept. Psychophys. **62**, 1312–1319.

Slis, I. H., and Cohen, A. (**1969a**). "On the complex regulating the voiced-voiceless distinction I," Lang Speech **12**, 80–102.

Slis, I. H., and Cohen, A. (**1969b**). "On the complex regulating the voiced-voiceless distinction II," Lang Speech **12**, 137–155.

Slis, I. H., and Van Heugten, M. (**1989**). "Voiced-voiceless distinction in Dutch fricatives," in *Linguistics in The Netherlands 1989*, edited by H. Bennis and A. Van Kemenade (Foris, Dordrecht, The Netherlands), pp. 123–132.

Stevenson, D. C. (**1979**). "Categorical perception and selective adaptation phenomena in speech," Ph.D. dissertation, University of Alberta, Edmonton, Canada.

Strange, W. (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (York Press, Baltimore, MD).

Van den Berg, R. J. H. (**1987**). "Effects of durational factors on the perception of voicing in Dutch two-obstruent sequences," J. Phonetics **15**, 259–271.

Van den Berg, R. J. H. (**1988**). "The perception of voicing in Dutch two-obstruent sequences: A comparison of synthetic and natural speech," J. Phonetics **16**, 171–180.

Van den Berg, R. J. H. (**1989**). "Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues," Speech Commun. **8**, 17–25.

Watson, I. (**1983**). "Cues to the voicing contrast: A survey," Cambridge Papers in Phonetics and Experimental Linguistics **2**, 1–34.

Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants[a)]

Catherine M. Siciliano,[b)] Andrew Faulkner, Stuart Rosen, and Katharine Mair
Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, UCL, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 14 December 2007; revised 8 December 2009; accepted 22 December 2009)

Simulations of monaural cochlear implants in normal hearing listeners have shown that the deleterious effects of upward spectral shifting on speech perception can be overcome with training. This study simulates bilateral stimulation with a unilateral spectral shift to investigate whether listeners can adapt to upward-shifted speech information presented together with contralateral unshifted information. A six-channel, dichotic, interleaved sine-carrier vocoder simulated a binaurally mismatched frequency-to-place map. Odd channels were presented to one ear with an upward frequency shift equivalent to 6 mm on the basilar membrane, while even channels were presented to the contralateral ear unshifted. In Experiment 1, listeners were trained for 5.3 h with either the binaurally mismatched processor or with just the shifted monaural bands. In Experiment 2, the duration of training was 10 h, and the trained condition alternated between those of Experiment 1. While listeners showed learning in both experiments, intelligibility with the binaurally mismatched processor never exceeded, intelligibility with just the three unshifted bands, suggesting that listeners did not benefit from combining the mismatched maps, even though there was clear scope to do so. Frequency-place map alignment may thus be of importance when optimizing bilateral devices of the type studied here.
 [DOI: 10.1121/1.3293002]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Sr, 43.66.Rq [KWG] Pages: 1645–1660

I. INTRODUCTION

For users of a cochlear implant (CI), the restoration of binaural hearing could provide a range of advantages including binaural squelch, the head-shadow effect, and increased stimulus redundancy, all of which may lead to better understanding of speech in noisy listening environments. Even though current CIs do not provide the fine structure information that contributes to binaural processing in normal hearing, a binaural advantage for speech perception has been reported in patients with bilateral cochlear implants (Dorman and Dahlstrom, 2004; Tyler *et al.*, 2005; Litovsky *et al.*, 2006; Wackym *et al.*, 2007; Tyler *et al.*, 2007) and for implants used in conjunction with acoustic hearing aids (HAs) (Iwaki *et al.*, 2004; Ching *et al.*, 2004; Hamzavi *et al.*, 2004; Ching *et al.*, 2005; Ching, 2005; Ching *et al.*, 2006).

While the potential benefits are manifold, the bilateral use of auditory devices raises new questions in CI frequency-place mapping. Cochlear implant electrode arrays are usually designed for an insertion depth of 25 mm into the typically 35 mm long cochlea. In many cases, however, the insertion achieved is shallower than this. Estimates based on *in vivo* computed tomography from 26 Nucleus-22 implant recipients showed insertion depths ranging from 11.9 to 25.9 mm, as well as considerable variation in cochlear length from 29.1 to 37.4 mm (Ketten *et al.*, 1998; Skinner *et al.*, 2002). The topography of the neural elements stimulated by the electrode contacts is not at present completely understood. The standard approach for estimating the effective characteristic frequency (CF) at each electrode contact has followed the frequency-to-place mapping of the organ of Corti established by Greenwood (1990). At the median insertion depth of 20 mm found by Ketten *et al.* (1998), the Greenwood map leads to an estimated CF of 1000 Hz. However, this map is not a realistic model for the CFs of CI electrodes placed near the modiolus, for which a spiral ganglion map seems more appropriate (Stakhovskaya *et al.*, 2007). A spiral ganglion frequency-to-place map assigns substantially lower CFs to a given electrode contact position than does Greenwood's organ of Corti map, especially for more apical electrode locations. It may therefore be important to consider proximity to the modiolar wall when estimating the effective CF of an electrode contact, which would require imaging of the cochlea and the electrode array for each subject.

Notwithstanding our incomplete knowledge of the effective CFs along an electrode array, the altered frequency-to-place mapping resulting from incomplete insertion of the implant electrode array and band-limited speech processing has demonstrable and sizable effects for monaural CIs (Skinner *et al.*, 2002). This has led to an ongoing debate about whether it is best to preserve tonotopic matching at the ex-

[a)] Portions of this work were presented at the 2006 Midwinter Meeting of the Association for Research in Otolaryngology, the 2007 Conference on Implantable Auditory Prostheses, and the 2007 British Society of Audiologists Short Papers Meeting on Experimental Studies of Hearing and Deafness.

[b)] Author to whom correspondence should be addressed. Present address: Audiology and Deafness Research Group, School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom. Electronic mail: sicilian@post.harvard.edu

pense of a frequency-shifted map that is maximally informative for speech. Earlier acute studies using vocoder-based simulations of cochlear implant speech processing suggested that a basalward frequency shift larger than 3 mm, using an organ of Corti map, leads to large decreases in speech intelligibility (Dorman *et al.*, 1997a; Shannon *et al.*, 1998). Analogous downward shifts of CI processor analysis filter frequencies have been shown to have a similar effect on vowel recognition by CI patients (Fu and Shannon, 1999). Yet, more recent studies with both noise-vocoded simulations (Rosen *et al.*, 1999; Fu *et al.*, 2005b; Faulkner *et al.*, 2006; Smith and Faulkner, 2006) and cochlear implant patients (Fu *et al.*, 2002; Svirsky *et al.*, 2004; Fu *et al.*, 2005a) have allowed listeners time to adapt to altered frequency-place maps and have shown improved performance after a period of training. For example, Rosen *et al.* (1999) trained normal hearing listeners with noise-vocoded simulations of a 6.5 mm basalward shift and found that word recognition had improved from floor to 30% after 3 h of training.

However, with bilateral implants or a cochlear implant in combination with a contralateral HA, the frequency-place maps in the two ears may be very different. For bilateral implants, the two electrodes may be inserted to different depths, resulting in different degrees of basalward shift. This may also interact with varying patterns of nerve survival in the two ears. In the case of CI+HA, the HA ear will retain the natural frequency-place map (albeit with a limited frequency range), while the CI ear is likely to be subject to basalward place shifting. To what extent are listeners able to adapt when presented with frequency-place maps that differ between the two ears?

Dorman and Dahlstrom (2004) reported a binaural advantage for speech perception in two bilateral implant patients who had different cochlear implants in each ear, which may support the hypothesis that information from mismatched frequency-to-place maps can be combined. Patients showed improvements of 32%–34% on HINT sentences with the addition of the second implant over performance with the better ear alone. However, the study only included two subjects, and the method for determining mismatch between the ears—pitch-ranking of electrodes—was inexact. The authors conceded that the degree of mismatch between the ears was unknown, and thus the question of adaptation to such a mapping remains open.

Evidence from dichotic listening experiments suggests that speech information presented in complement across the two ears is easily integrated. Broadbent and Ladefoged (1957) showed that listeners presented with the F1 and F2 of a /da/ syllable separately to opposite ears perceived the syllable as /da/. This process, later termed *spectral fusion* by Cutting (1976), is robust to differences in level and fundamental frequency, but not relative onset time; the majority of listeners appear able to integrate (tonotopically matched) acoustic cues efficiently (Rubin *et al.*, 1992).

To the extent that the information delivered to each ear with bilateral CIs is in complement, a binaural advantage for speech perception should be expected. However, this might be achieved through various possible mechanisms. For example, the binaural combination of information is likely to be relatively easy if there are similar basalward shifts to each ear. Alternatively, if listeners were able to adapt to different degrees of basalward shift in each ear, then they might still be able to combine spectral cues from the two ears to show a binaural speech advantage: information that may have been perceived as conflicting initially may eventually be perceived as complementary after experience with the implants. However, many bilateral CI users do not show a binaural advantage for speech over the best ear alone for speech in quiet (Tyler *et al.*, 2007) or for speech in noise in the absence of directional differences (van Hoesel *et al.*, 2002; Wackym *et al.*, 2007). A recent study suggests that better-ear listening is especially apparent for bilateral CI users who show asymmetrical monaural speech scores in each ear (Mosnier *et al.*, 2009). It thus seems plausible that misaligned frequency-place maps between each ear may underpin a lack of binaural advantage for speech shown in some bilateral CI users.

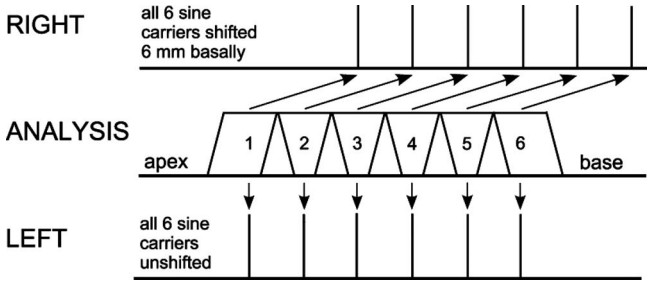


FIG. 1. Speech processing depicting sine-vocoded simulations of bilateral cochlear implants. Here, all analysis channels are presented to both ears, with the sine carriers at the right ear being upwardly shifted by an equivalent of 6 mm basilar membrane difference. In this configuration, the spectral information in each analysis band is delivered to different places in each ear. Further, the carrier frequencies for bands 1–4 in the right ear are the same as the carrier frequencies for bands 3–6 in the left ear so that there is an inter-aural conflict at each of these places.

A. Considerations for simulation methods sensitive to binaural advantage

Compared to monaural stimulation with cochlear implants, bilateral stimulation introduces complicating factors that may lead to decrements in speech perception. Consider the simplest case where contiguous frequency bands are vocoded and presented to each ear with an upward shift to one ear only, as in the bilateral six-channel sine vocoder illustrated in Fig. 1. Two distinct aspects of mapping conflict are evident. First, for each analysis band, the same spectral content is delivered to different tonotopic locations in each ear. Second, the carrier frequencies for bands 1–4 in the shifted ear are tonotopically aligned to the carriers for bands 3–6 in the unshifted ear so that the spectral information carried at each of these corresponding places is different in the two ears. The second of these conflicts may in addition lead to central masking (Mills *et al.*, 1996). A related but conceptually distinct global consequence of a mismatch of frequency mapping between the ears is that the central processing of spectral speech features cannot integrate information from the two ears without reference to an ear-specific transformation of the peripheral excitation pattern which needs to be learned.

A bilateral implant simulation of the type shown in Fig. 1, while representative of the information presented to an implant user, is not ideal for the study of adaptation to interaural mismatches of frequency mapping in normally hearing listeners. The primary difficulty with such a simulation is the redundancy of information between the shifted and unshifted ears. For tonotopically aligned speech in quiet, a binaural advantage can only be expected when there is some nonredundant information in each ear that can be exploited. The binaural advantage shown in some users of bilateral cochlear implants may arise from inter-aural variations in nerve survival or differences in spatial specificity at each ear that result in a sparcity of redundant information in each ear. However, for the normal hearing listener the signals will be completely redundant, which is especially problematic in a study where the intent is to encourage adaptation to a spectral shift to one ear only, because the information that can be gained by such adaptation is minimal. A second limitation of a bilateral implant simulation that employes the same contiguous analysis bands for both ears is that the local conflicts noted above are confounded with the global conflict that arises from the interpretation of two mismatched excitation patterns.

This redundancy can be eliminated through a configuration related to so-called *zipper processing*, which was initially proposed as an alternative processing strategy for bilateral CIs. Here only alternate electrodes at each ear were excited in an interleaved fashion with the aim of reducing interaction between electrodes (Lawson *et al.*, 2000). Data from pitch-ranking of electrodes across the ears were used to ensure, as far as possible, that the two ears were fitted without a significant mismatch of frequency-to-place mapping. The technique did not lead to improved speech scores compared to monaural fittings with the same total number of channels, but nor did it lead to decrements in speech intelligibility. While "zipper processing" has not been implemented clinically and is thus not representative of typical bilateral CI processing, the approach does provide a configuration that reduces redundancy and local conflicts of mapping between the two ears. It was thus ideal for the consideration of mismatched maps in normally hearing listeners in the absence of the confounding effects outlined above.

All speech processing was based on sine-excited vocoding (Loizou *et al.*, 2003). Sine carriers were chosen primarily to constrain the excitation patterns to avoid overlap of excitation in the two ears, which would be likely at the edges of noise-band carriers. The zipper-like (Lawson *et al.*, 2000) interleaved processing was chosen because it maximizes the potential advantage from the binaural combination of information between the two ears. In tests of dichotic speech perception in noise, Loizou *et al.* (2003) showed that for dichotically presented (unshifted) noise-vocoded speech, intelligibility was higher (and thus presumably binaural information more effectively combined) when channels were interleaved as here, rather than split according to low and high frequencies. Similar results were found for quiet speech in users of bilateral cochlear implants (Mani *et al.*, 2004).

The extent of mismatch between the ears was large at 6 mm to simulate the effect of a shallowly implanted electrode in one of two stimulated ears. Six channels were used, interleaved as three channels to each ear, to minimize the possibility of ceiling performance in unshifted conditions and at the same time to provide a shifted signal with sufficient information to be learnable. In relation to ceiling limits, it has been widely found that an unshifted three-channel vocoder leads to substantially poorer scores for most speech materials than an unshifted six-channel vocoder (Dorman *et al.*, 1997b; Fishman *et al.*, 1997; Loizou *et al.*, 1999). Thus, any improvements gained through adaptation to a mismatch between the two ears should be easily detectable. A pilot study showed that unshifted conditions were prone to ceiling effects with some easier speech materials, so the use of more than six channels in quiet was not explored (Faulkner, 2006). In relation to learnability, a 6.5 mm shifted four-band processor has been shown to allow fairly rapid adaptation given training (Rosen *et al.*, 1999).

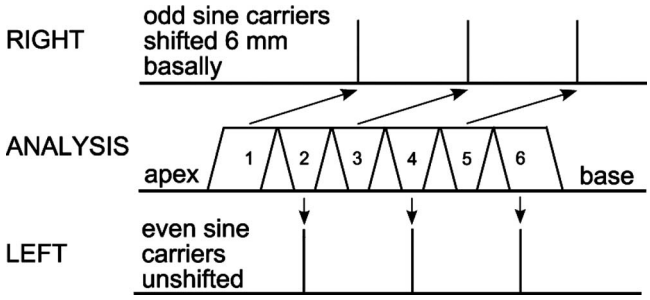


FIG. 2. Speech processing producing a binaurally mismatched frequency-to-place map. The middle panel represents the six spectral analysis bands covering the frequency range 200–5000 Hz. Temporal envelope information from channels 1, 3, and 5 is presented to the right ear (see top panel) imposed on sinusoidal carriers whose frequencies are shifted upward compared to the analysis channels to an extent equivalent to a 6 mm basalward shift on the basilar membrane. Temporal envelope information from channels 2, 4, and 6 is presented to the left ear (bottom panel) imposed on sinusoidal carriers that match the center frequencies of these three analysis bands.

Figure 2 shows a schematic representation of the binaurally mismatched processor that was employed. From low frequency (apical end) to high frequency (basal end), odd-numbered bands 1, 3, and 5 were presented to the right ear with the equivalent of a 6 mm basalward basilar membrane shift (assuming a 35 mm cochlea). Even-numbered bands 2, 4, and 6 were presented to the left ear without a shift.

B. Overview of research questions

In Experiment 1, we trained one group of listeners with a binaurally mismatched processor over eight 40 min training sessions, testing the extent of adaptation before, during, and after training. It was hypothesized that the presence of moderately intelligible unshifted frequency components may hinder adaptation to the shifted frequency map and thus to the binaurally mismatched processor. Consequently we also trained and tested a second group of listeners with just the shifted components to examine whether adaptation to the binaurally mismatched map could be facilitated by training with the shifted processor alone in the absence of the conflict between the two maps. In Experiment 2, we looked at how listeners adapt to the binaurally mismatched map over an

TABLE I. Conditions for Experiment 1.

Condition	Abbreviation	Component bands and shift	
		Right	Left
Six dichotic unshifted	6DU	1, 3, 5	2, 4, 6
Six dichotic odd shifted	6DOS	1, 3, 5 → 6 mm	2, 4, 6
Six odd shifted	6OS	1, 3, 5 → 6 mm 2, 4, 6	1, 3, 5 → 6 mm 2, 4, 6
Three even unshifted	3EU		2, 4, 6
Three odd shifted	3OS	1, 3, 5 → 6 mm	
Three odd unshifted	3OU	1, 3, 5	

extended course of training. We doubled the length of training and alternated training conditions from that of the two groups from Experiment 1 to see whether adaptation to binaurally mismatched frequency-place maps required a longer period and/or different type of training.

Perceptual adaptation to binaurally mismatched frequency-place maps has not been previously explored. If listeners do not learn to benefit from the binaurally mismatched maps, then this may reflect a constraint on plasticity for speech perception that could have implications for bilateral cochlear implant fittings. Alternative frequency-place mappings that avoid such binaural mismatches might be more optimal for speech recognition. If, however, a spectral mismatch can be learned, then the circumstances which facilitate this adaptation, such as specialized training techniques, need to be clarified.

II. EXPERIMENT 1

A. Method

1. Subjects

Twelve normally hearing speakers of British English took part, and each was paid for his or her participation. All had pure-tone audiometric thresholds better than 20 dB hearing level (HL) at octave frequencies from 250 to 8000 Hz. Ethical approval for this study was granted by the UCL/UCLH Joint Committee on the Ethics of Human Research.

2. Test conditions

Table I summarizes the six conditions tested in Experiment 1. These conditions were designed to assess what aspects of the signal were involved in any learning that took place. Test conditions were composed of monotic, dichotic, and diotic combinations of the spectral components outlined in Fig. 2 for the binaurally mismatched processor. Condition names were coded by number of channels, dichotic presentation, even- or odd-numbered channels, and shifted or unshifted presentation. When basalward shift was applied, it was always to the odd-numbered channels. Even-numbered channels were always unshifted.

The six dichotic unshifted (6DU) condition served as a control to assess maximal intelligibility. Here, unshifted odd- and even-numbered channels were interleaved between the ears. A pilot study indicated that this dichotic unshifted processor yielded intelligibility equivalent to a six-channel diotic processor. The main experimental condition (Fig. 2) was the six dichotic odd shifted (6DOS) condition. Here, odd-numbered bands were shifted an equivalent of 6 mm basally and presented to the right ear, while even-numbered bands were presented to the left ear without a shift. This binaurally mismatched processor was also used for training in one group of subjects. The six odd shifted (6OS) condition comprised the same six envelope-modulated carriers as the 6DOS condition, but was instead diotic: all bands were presented to both ears. In contrast to 6DOS, 6OS lacked any cue from ear of presentation to those carrier bands that were shifted. If performance here differed markedly from the 6DOS condition, this would suggest that listeners were learning to attune to information in an ear-specific manner. Conversely, if there was no difference, then we would infer that learning was occurring on the basis of the carrier frequencies. In the three even unshifted (3EU) condition, even-numbered channels alone were presented to the left ear unshifted. This was equivalent to the unshifted components of the 6DOS processor. In the three odd shifted (3OS) condition, odd-numbered channels were presented alone to the right ear with a 6 mm basalward shift. This comprised the shifted components of the 6DOS processor. This processor was also used for training in a second group of subjects. In the three odd unshifted (3OU) condition, odd-numbered channels were presented to the right ear unshifted. This condition was only tested for the group trained with the 3OS processor and allowed a comparison of the information provided by the odd- and even-numbered bands in the absence of shifting.

3. Signal processing

Center and crossover frequencies for the analysis and output filters were calculated using Greenwood's equation and its inverse, relating distance x (in mm) from the apex along the basilar membrane to characteristic frequency (in Hz). The assumed cochlear length was 35 mm (Greenwood, 1990).

$$\text{frequency} = 165.4(10^{0.06x} - 1),$$

$$x = \frac{1}{0.06}\log\left(\frac{\text{frequency}}{165.4} + 1\right).$$

The amplitude envelope of each band was extracted with an analysis filter, full-wave rectification, and a smoothing filter. The envelope was then multiplied by a sinusoid with frequency matching the center frequency of the band (or shifted equivalent). Finally, the requisite bands were summed and presented to the left and/or right ears as determined by pro-

TABLE II. Analysis band cutoff and carrier frequencies for each band in the unshifted (6DU, 3EU, and 3OU) and odd-band shifted (6DOS, OS, and 3OS) conditions.

Band	Analysis band cutoff (Hz)		Analysis band center frequency (Hz)	Carrier frequency (Hz)	
	Lower	Upper		Unshifted	Odd-shifted
1	200	403	290	290	878
2	403	718	543	543	543
3	718	1208	936	936	2359
4	1208	1971	1547	1547	1547
5	1971	3157	2498	2498	5937
6	3157	5000	3977	3977	3977

cessor condition. Table II shows input and output center frequencies as well as filter cutoffs. All processor conditions used the same six analysis filters and sine carriers at either the shifted or unshifted center frequencies of these analysis filters.

A real-time implementation of the vocoder processor was used for live training, while offline processing of the test material was implemented in MATLAB. This ensured identical repetition of test materials. Offline processing was executed at a 44.1 kHz sampling rate. Analysis bands were determined by a serial implementation of high-pass and low-pass third-order Butterworth IIR filters. Adjacent filter responses crossed at 3 dB down from the peak of the pass-band. Envelope smoothing used second-order low-pass Butterworth filters with a 32 Hz cutoff. Real time processing was implemented using the Aladdin Interactive DSP Workbench (Hitech Development AB, Sverige) and ran on a DSP card (Loughborough Sound Images TMSC31, Loughborough, UK). The computational power of the DSP was limited so the sampling rate was restricted to 16 kHz, and elliptical rather than Butterworth filter designs were used with the same 3 dB crossover frequencies as for the offline processing. Analysis filters consisted of fourth-order band-pass designs, while third-order low-pass filters were used for envelope smoothing.

In both testing and training, an equal loudness correction was applied to each of the shifted bands to preserve relative loudness across the spectra of unshifted and shifted speech. The correction was set to half the difference (in dB) between the minimal audible field threshold of the analysis filter and that at the center frequency of the shifted output filter. Minimal audible field values were taken from Robinson and Dadson (1956) and interpolated using a cubic spline fit to log frequency.

4. Training

Subjects were trained with connected discourse tracking (CDT) (De Filippo and Scott, 1978). In this method, the experimenter reads successive phrases from a text to the subject, who then repeats back what he or she heard. This allows the listener to acclimate to the spectrally distorted speech while engaging in a communication task that is similar to a conversation. The number of words repeated back correctly per minute provides a measure of progress throughout training. CDT has been shown to be an effective training method for spectrally shifted speech (Rosen *et al.*, 1999; Faulkner *et al.*, 2006).

The talker for the CDT portion of this experiment was the first author, C.S. Although she is a native speaker of a north-eastern dialect of American English, she had been living in the U.K. for 5 years at the time of testing and has been judged to have an accent similar to Standard Southern British English. CS's speech was not used for any of the testing. The talker read from the text in short phrases, and the listener repeated back what he or she heard. If the listener's response matched what the talker had said, the talker would move on to the next phrase. Otherwise the phrase was repeated. If after the third presentation, the listener could still not reproduce the phrase, the listener was presented the phrase as unprocessed speech (to the left ear only). Texts for CDT were chosen from the Heinemann Guided Readers series (elementary level). These texts are designed for learners of English as a second language and make use of controlled vocabulary and syntactic complexity. During training, the talker and subject were situated in adjacent sound-treated rooms. The room had a double-glazed window that enabled auditory-visual (AV) training. During auditory-only training, the window was blinded. A constant pink masking noise at 45 dBA was played in the listener's room to mask any speech from the talker that might be transmitted through the wall and window. The talker heard the listeners' responses over an intercom, and no attempt was made to prevent the listener hearing their own unprocessed voice when responding. Of the 12 subjects, 6 were trained with the 6DOS processor (6DOS-trained group), and the remaining 6 were trained with the 3OS processor (3OS-trained group).

5. Test materials

a. Sentence perception The IEEE/Harvard sentence lists (Rothauser *et al.*, 1969) were used, which have very little contextual information. Digital recordings of the sentences were from one male and one female talker of British English (16 bit, 48 kHz downsampled to 44.1 kHz). The 72 lists in the set each contained ten sentences with five keywords in each sentence. The first 36 lists were designated for the female talker, and the remaining 36 lists were designated for the male. A subset of 32 lists from each talker was used for the 6DOS-trained group, who were tested with fewer conditions. For each test session, two lists per condition were chosen from each talker set in a pseudo-random manner. No list appeared more than twice in the same condition across all of the subjects, and subjects never heard a list more than once. The subject was asked to repeat back to the experimenter as many words as he or she could, and no feedback was given. Words were counted correct when the word root was repeated correctly.

b. Vowel identification Vowel identification was included as a measure of the contribution of spectral cues in each condition and also as a source of both confusion and recognition accuracy data. If vowel confusions in the presence of spectral shift become more similar to those for unshifted processors after training, then this would reflect adaptation to the shifted speech. The task also allowed for the

TABLE III. Sequence of training and testing conditions for Experiment 1.

Session	Training	Processor	Testing
1	5 min audio visual (AV), 5 min auditory alone (AA)	6DU	*Familiarization:* Unprocessed vowels; *Pre-test*: IEEE sentences [2 lists × 5(6) conditions × 2 talkers], bVd identification [5 tokens × 5(6) conditions × 2 talkers]
2—5	5 min AV, 35 min AA	6DOS or 3OS	None
6	Familiarization if not immediately after session 5 (see text)		*Mid-test*: IEEE sentences [2 lists × 5(6) conditions × 2 talkers], bVd identification [5 tokens × 5(6) conditions × 2 talkers]
7—10	5 min AV, 35 min AA	6DOS or 3OS	None
11	Familiarization if not immediately after session 10 (see text)		*Post-test*: IEEE sentences [2 lists × 5(6) conditions × 2 talkers], bVd identification [5 tokens × 5(6) conditions × 2 talkers], IHR/BKB sentences (2 lists × 6 conditions × 2 talkers) (3OS-trained group only)

comparison of the intelligibility of the information contained in the unshifted odd and even subsets of channels, which was important for demonstrating that the information in these subsets of channels was not entirely redundant, and that there was room for improvement beyond the three channel unshifted speech. Further, since vowels can be described primarily by their first and second formants, vowel confusions can be easily mapped in terms of the relationship between the expected vowel and the output at the sine-carrier frequency of a given formant.

Nine b-vowel-d words in the carrier sentence "Say bVd again" were recorded by a male and female speaker of British English in anechoic conditions at a 48 kHz sampling rate and subsequently downsampled to 44.1 kHz. The male but not the female talker was the same as for the sentence test. Five tokens of each bVd word were recorded from each talker so that in an individual test of one talker in a given condition, there were 45 items. Vowels were restricted to monophthongs of similar duration so that listeners would need to rely on spectral cues for identification: /æ/ (bad), /ɑː/ (bard), /iː/ (bead), /e/ (bed), /ɜː/ (bird), /ɔː/ (board), /ɒ/ (bod), /uː/ (booed), and /ʌ/ (bud). A grid with all nine words appeared, and the subject clicked with the computer mouse on the button displaying the word they perceived. The vowels were represented on the buttons in the orthographic form given above. Before testing, the subject was given a practice session in which the vowel material was presented unprocessed, with a single token for each vowel and each talker. This enabled the subjects to familiarize themselves with the software and the task.

6. Procedure

Subjects were tested before training commenced, halfway through training, and at the end of training. All subjects completed the entire cycle of training and testing within a maximum of 2 weeks, with no more than a 2 day gap between successive sessions. Sentence and vowel test presentation was counterbalanced across the group. Within each test, stimuli were pseudo-randomized by block of condition and talker. Table III presents the sequence of training and testing for Experiment 1.

In the first session, subjects were acclimatized to unshifted sine-vocoding with a 10 min block of CDT with the DU processor prior to the pre-training test session. Previous experiments have demonstrated that listeners require a short period of acclimatization before they can reliably perceive speech through unshifted vocoders with a limited number of channels (Davis *et al.*, 2005). As in subsequent CDT training sessions, the first 5 min of the familiarization block were AV, while the remaining 5 min were auditory alone (AA). Following the pretest, subjects were trained with CDT in four 40 min training sessions with either the dichotic odd shifted speech (6DOS-trained group) or the monaural three shifted channel speech (3OS-trained group). Subjects were tested after the fourth training session. If the testing session did not take place immediately following a training session, then the experimenter administered a 10 min (five AV, five AA) CDT block with the 6DOS or 3OS processor, which was not counted toward the total hours of training. Following this mid-training test session, subjects underwent four more 40 min training sessions with the same processor as in the first four sessions, and then completed the post-training testing.

Testing took place in a sound-treated room with presentation of the processed speech over Sennheiser HD280 headphones. The level was set by the experimenter to a comfortable listening level, and this level was used by all participants.

B. Results

Test data were analyzed using repeated measures analysis of variance (ANOVA), with within subject factors of test session, condition, and talker, and where relevant, a between subjects factor of processor used in training. Hyunh–Feldt epsilon corrections were applied to all F tests for factors with more than one degree of freedom. Hyunh–Feldt adjusted degrees of freedom have been rounded to the nearest integral value, and the significance criterion was $p=0.05$. *A priori* hypotheses were tested using planned contrasts, and *post-hoc* testing was carried out using Bonferroni-adjusted paired comparisons. Data were typically pooled by talker, unless there was a significant interaction between talker and processor that could not be attributed to floor effects with the shifted (3OS) condition.

1. IEEE sentence perception

a. 6DOS-trained group Keywords correct for the IEEE sentence test across training sessions for the two talkers combined are shown in the left-hand panel of Fig. 3. For the

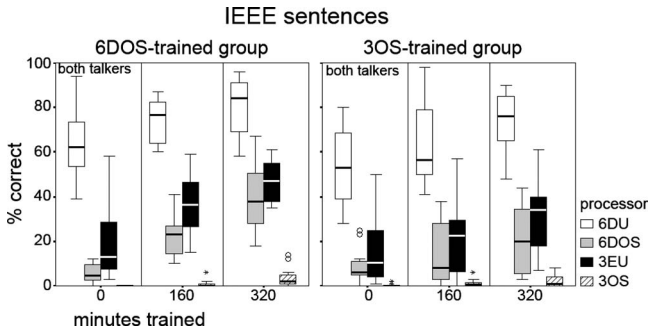


FIG. 3. Experiment 1 box-and-whisker plots of IEEE sentence scores as a function of training time. The box shows interquartile range over six subjects, the bar shows the median, and the whiskers show the complete range excluding any outlying values (shown as open circles or asterisks). Plots are shown for the two talkers combined. A vertical reference line separates each training time period. Scores for the 6DOS-trained group are in the left-hand panel, while scores for the 3OS-trained group are in the right panel. The plots show the four main conditions: 6DU, 6DOS, 3EU, and 3OS.

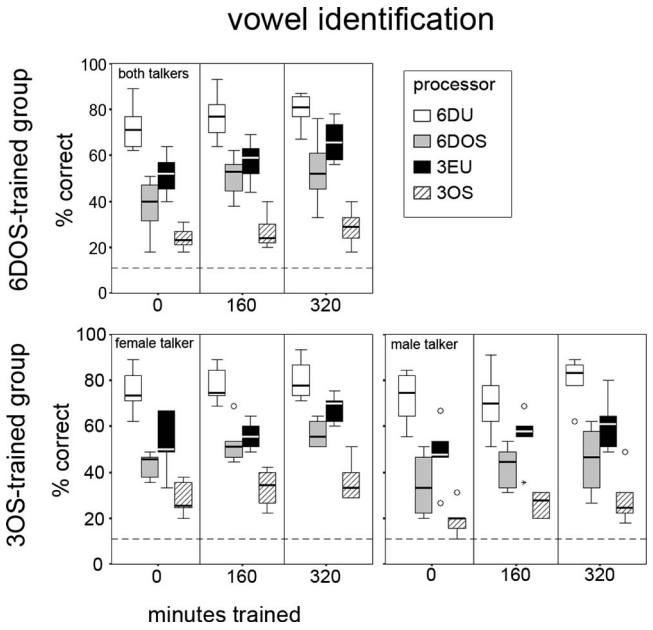


FIG. 4. Experiment 1 vowel identification scores as a function of training time. Scores for the 6DOS-trained group are in the top panel, while scores for the 3OS-trained group are in the bottom two panels. The 3OS-trained group showed a significant talker by processor interaction, so plots for each of the two talkers are given separately. The vertical reference line separates the training time periods, and the horizontal reference line indicates chance performance, which was 11.1%.

6DOS-trained group, performance with the 3OS processor remained close to floor throughout training, while intelligibility with all other processors tended to increase. This was indicated by a significant training session with processor interaction [$F(8,40)=445$, $p<0.001$]. Scores have been pooled by talker, although they were slightly better with the male talker in all but the 3OS condition, which is reflected by the significant interaction of talker with processor [$F(3,17)=6.57$, $p=0.003$]. When the data were reanalyzed excluding the 3OS condition, these interactions were no longer significant, and significant main effects were evident for number of training sessions [$F(2,10)=71.5$, $p<0.001$] and processor [$F(2,9)=146$, $p<0.001$].

Post-hoc testing on the post-training sentence scores revealed three key findings. First, performance with the three unshifted channels (3EU) was significantly worse than with the dichotic unshifted (6DU) condition ($p=0.001$), which is a clear indication that there was room for improvement with the additional channels in the shifted ear. Second, there was no significant difference between the 6DOS condition and the 3EU condition, which indicates that subjects did not show a binaural advantage. If they had, we would expect to see performance with this processor exceeding that in the 3EU condition, as was found for the 6DU condition.

b. 3OS-trained group Scores for IEEE sentence recognition for the 3OS-trained group are shown in the right-hand panel of Fig. 3. A repeated measures ANOVA showed significant main effects of number of training sessions [$F(2,10)=36.0$, $p<0.001$] and processor [$F(3,16)=85.9$, $p<0.001$], but there was no significant training session with processor interaction, indicating significant improvement with training that was relatively similar in all conditions, including 3OS.

Post-hoc comparisons of post-training sentence scores for the 3OS-trained group were similar to the 6DOS-trained group across conditions. Performance with 6DU was significantly better than 3EU ($p<0.001$), while there was no significant difference in intelligibility between 6DOS and 6EU. Thus while the intelligibility scores for this group were generally lower and improvements with training smaller than for the 6DOS-trained group, the pattern between conditions is similar. Subjects adapted to the vocoder processing, but there is no evidence of a binaural advantage with the mismatched frequency-place maps.

2. Vowel identification

a. 6DOS-trained group Vowel identification scores for the two talkers combined are summarized in the top panel of Fig. 4. The overall pattern of results is similar to that seen for sentences. There were significant main effects of number of training sessions [$F(2,10)=37.4$, $p<0.001$] and processor [$F(4,20)=102$, $p<0.001$] and a significant training sessions with processor interaction [$F(8,40)=89.6$, $p=0.005$]. The lower bound of the 95% confidence interval was above chance level performance (11%) at all test points and for all conditions, confirming that vowel recognition scores were significantly greater than chance.

Importantly, *post-hoc* testing showed significant improvements in all conditions between the pre-test and post-test sessions ($p<0.05$), including 3OS, and a planned contrast of training sessions within the 3OS condition also showed significant improvement after training in this condition ($p=0.01$). While the significant interaction of training session with processor may indicate that improvement with the 3OS condition was smaller than the improvement in the other conditions, this is nonetheless evidence of learning in this condition. Despite the improvement in the 3OS condition, again there was no significant difference between 3EU and 6DOS after training, even though performance with 3EU was significantly worse than with 6DU ($p=0.02$). Thus, while there was room for improvement by the addition of the odd channels (3EU<6DU), and there was evidence of learning

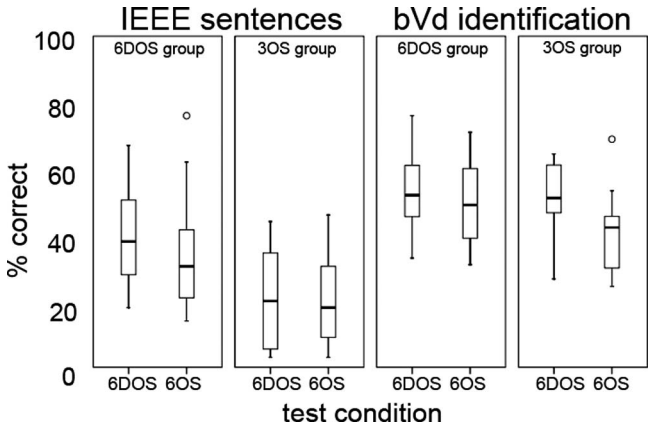


FIG. 5. Comparison of dichotic (6DOS) versus diotic (6OS) presentation of mismatched frequency-place maps within each stimulus type and training group from Experiment 1. The IEEE scores are in the left two panels (6DOS and 3OS groups, respectively), while the bVd identification scores are in the two panels on the right (6DOS and 3OS groups, respectively).

when these channels were shifted, subjects did not show a binaural advantage with the mismatched frequency-place maps.

b. 3OS-trained group The bottom panels of Fig. 4 show vowel identification scores for the 3OS-trained group. A repeated measures ANOVA showed a significant talker with processor interaction [$F(5,25)=3.58$, ($p=0.014$)], so the boxplots and subsequent analyses are given for each talker individually. In Bonferroni-corrected comparisons of talker within each condition, intelligibility with the female talker was significantly better than with the male only for conditions where spectral shifting was present, i.e., 6DOS, 6OS, and 3OS. In unshifted conditions, there was no significant difference in intelligibility between the talkers. This may indicate a gender-specific adaptation to the shifted speech since all experience with shifted speech outside of testing was also with a female talker. More evidence for this comes from a planned contrast of test sessions within the 3OS condition, which showed significant improvement with training only for the female talker ($p=0.009$).

For both talkers, the lower 95% confidence limit was above chance for all test conditions and at all test points. The ANOVAs also showed significant main effects of session [female: $F(2,10)=21.2$, ($p<0.001$); male: $F(2,10)=10.1$, ($p=0.004$)] and processor [female: $F(5,25)=45.7$, ($p<0.001$); male: $F(4,19)=33.7$, ($p<0.001$)], but no significant session by processor interaction for either talker. Like the IEEE sentences for this group, then, this analysis suggests that there were significant and similar levels of learning for this group in all conditions, including 3OS.

3. Comparison of dichotic versus diotic presentation

Mismatched maps were compared under dichotic and diotic conditions to test the hypothesis that subjects had learned to ignore information from the shifted ear. Boxplots comparing post-training intelligibility with the dichotic and diotic presentations of the mismatched maps are presented in Fig. 5. The left two panels give scores for IEEE sentences, and the right two panels for bVd identification. The results are presented for the two talkers combined. The scores were entered into a repeated measures ANOVA. For the 6DOS-trained group, there was no significant difference between dichotic (6DOS) versus diotic (6OS) presentation for sentences or vowels. This suggests that if this group had adapted by suppressing or ignoring information, this could only have been on the basis of carrier frequency rather than ear of input. By contrast in the 3OS-trained group, Bonferroni-corrected comparisons showed that post-training vowel identification was significantly worse for the diotic condition than the dichotic condition ($p=0.04$). This suggests that subjects in the 3OS-trained group were still relying on cues to ear presentation when listening to the mismatched maps since intelligibility was decreased when this cue was removed.

4. Comparison of unshifted odd and even channels

Intelligibility with unshifted even and odd channels (3EU and 3OU) was compared for the 3OS-trained group to confirm that the speech information in each subset of channels was not redundant. This is important for demonstrating that there was speech information that could be gained through the binaural combination of odd and even channels. The post-training results are summarized in the left (IEEE sentences) and right (vowel identification) panels of Fig. 6.

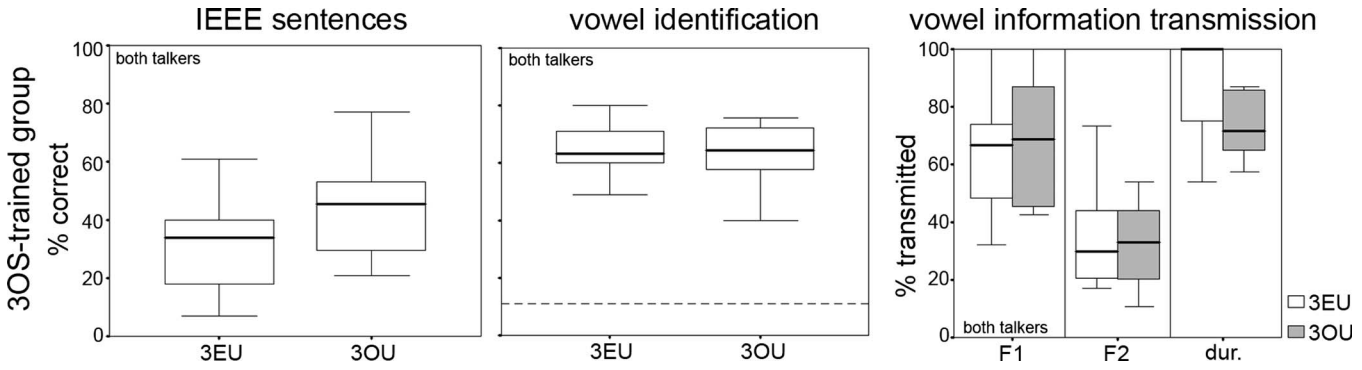


FIG. 6. Comparison of post-training scores for the unshifted even (3EU) and odd (3OU) channels for the 3OS-trained group of Experiment 1. The left panel shows IEEE sentence scores, the middle panel shows bVd identification, and the right panel shows percent of vowel feature information (F1, F2, and duration) transmitted. The results are shown for the two talkers combined.

TABLE IV. Feature definitions for the F1, F2, and duration features for each of the nine vowels. F1 was binary (open, close); F2 was tertiary (front, back, central); and duration was binary (long, short).

Feature	æ	ɑː	iː	e	ɜː	ɔː	ɒ	uː	ʌ
F1	o	o	c	o	c	o	o	c	o
F2	f	b	f	f	c	b	b	b	b
Duration	l	l	l	s	l	l	s	l	s

Surprisingly, intelligibility with the IEEE sentences was significantly better with the 3OU channels than the 3EU channels after training ($p=0.007$).

For vowel recognition, the ANOVA revealed no significant difference between these conditions for overall vowel recognition accuracy. The post-training vowel confusions for 3EU and 3OU were also analyzed in terms of information transfer (Miller and Nicely, 1955). Features of F1, F2, and duration were analyzed according to the feature definitions set out in Table IV. The results for the two talkers combined are summarized in the right-hand panel of Fig. 6. A repeated measures ANOVA revealed no significant difference between 3EU and 3OU in terms of information transfer for any of the three features. However, an χ^2 analysis showed that vowel confusions differed significantly between the two processors ($p<0.0001$). Hence there are clearly differences of detail between the vowel features conveyed by odd compared to the even bands, but no evidence of a difference across the set of vowels in the level of feature information provided by these two sets of bands.

These detailed differences in vowel confusions probably reflect the complementary representation of spectral information in each of the subsets of channels and also the sparse spectral coding which presents vowel formant energy only at the discrete carrier frequencies of the sine vocoder. For example, the /i:/ vowel was almost always incorrectly recognized as /u:/ in the odd unshifted condition (3OU) (87% of the time), but was recognized fairly accurately in the even unshifted condition (3EU) (77% correct). The F2 of this vowel for the female talker was around 2850 Hz, but the sine carrier for the analysis filter encompassing F2 was 2498 Hz, so F2 energy was lowered in frequency which could account for /i:/ being recognized as /u:/. For the /u:/ vowel, both F1 (319 Hz) and F2 (1980 Hz) frequencies fell within the analysis filters of odd-numbered bands. As would be expected, this vowel was recognized accurately in the odd channel condition (3OU) (85% correct), but in the even channel condition (3EU), accuracy was low (28%), and there was no consistent pattern in confusions since energy from formant frequencies was largely absent in this condition for this vowel. Taken together with the finding that 6DU was always better than 3EU in this experiment, there is thus clear evidence that the lack of binaural advantage shown with the binaurally mismatched processor (6DOS) compared to the even unshifted channels (3EU) is not attributable to a redundancy of speech information in the even and odd bands.

5. Vowel information transmission analysis

A further information transmission analysis of the post-training vowel confusions was performed to examine in more detail the information provided in the different processor conditions. The results for Experiment 1 are summarized in the left and middle panels of Fig. 7. Repeated measures ANOVAs for each training group were computed on the percentage of information transferred, with within-subjects factors of processor (6DOS, 3EU, and 3OS only), feature (F1, F2, and duration), and talker. For each group, there was a significant interaction of processor with feature [6DOS-trained group: $F(4,20)=3.17$, $p=0.036$; 3OS-trained group: $F(4,20)=12$, $p<0.001$]. Despite the equivalent post-training intelligibility scores for vowel identification, Bonferroni-corrected comparisons of processors within each feature revealed that the information transmitted with the 6DOS processor was significantly less than that with the unshifted 3EU processor, though this depended on feature and training group. For the F1 feature, this reduction in information was only marginally significant for the 6DOS-trained group ($p=0.048$), but was more significant for the 3OS-trained group ($p=0.008$). In fact, for the 3OS-trained group, there was no significant difference in F1 information for the 6DOS and 3OS conditions. Less F2 information was transmitted for 6DOS than 3EU for the 6DOS-trained group only ($p=0.03$), while less duration information was transmitted for the 3OS-trained group only ($p=0.025$). Not surprisingly, the information transmitted with the 3OS processor was always significantly worse than that with 3EU. This pattern of results suggests that subjects were not simply ignoring the shifted information in the 6DOS condition, and that training condition impacted the binaural combination of information.

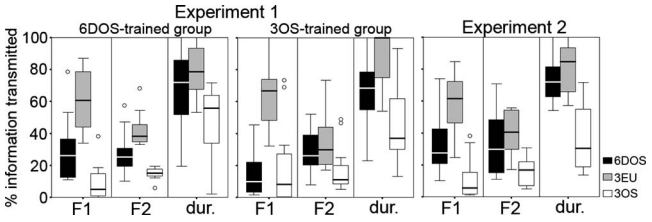


FIG. 7. Percent of vowel feature information (F1, F2, and duration) transmitted for the 6DOS, 3EU, and 3OS conditions. The results are shown for the two talkers combined. Scores are broken down by experimental group: the left panel shows Experiment 1 6DOS-trained group, the middle panel Experiment 1 3OS-trained group, and the right panel scores from Experiment 2.

C. Discussion

For both groups of subjects, performance with 6DU exceeded performance with 3EU in all but one subtest. Moreover, for the 3OS-trained group, the 3OU channels showed higher IEEE scores than the 3EU channels after training, and vowel confusions differed significantly with these processors. This is clear evidence that the information conveyed by the odd channels was not redundant in the presence of the even channels, which suggests that there was room for improvement if listeners could learn to use unshifted and shifted channels together for binaural advantage. However, neither group of subjects showed a binaural advantage for the binaurally mismatched processor since post-training intelligibility with the 6DOS processor never exceeded that with the 3EU processor for any of the speech materials.

Post-training intelligibility was similar in the 6DOS and 3EU conditions, which is consistent with the view that subjects had adapted to 6DOS by attending to the "better ear"—i.e., the unshifted frequencies. However, the information transfer analysis revealed that for both groups of subjects, vowel information transferred with the binaural 6DOS condition was significantly worse for some features than that for the monaural 3EU condition, which suggests a more complicated listening strategy. The binaural combination of shifted and unshifted channels may have led to a fused but "incorrect" percept—a process referred to as "psychoacoustic fusion" by Cutting (1976). It is also possible that the combination of shifted and unshifted channels was somehow distracting and thus led to more errors. The different pattern of results in the two groups suggests that the training condition affected the binaural combination of information. For the F1 feature, explicit training with the 6DOS condition seems to have mitigated the deficit with 6DOS compared to 3EU since for this group the difference was only marginally significant. By contrast, explicit training with 3OS seems to have been more beneficial for F2 information since only the 6DOS-trained group showed significantly worse F2 information transmission with 6DOS than 3EU.

Notwithstanding these apparent training condition effects, it seems clear that neither group of subjects learned to combine the shifted and unshifted channels in a way that led to a binaural advantage for speech over the time course in Experiment 1. It is possible that the period of learning required to show a binaural advantage for mismatched frequency-place maps may be longer than that needed for simple monaural basalward shifts. The improvements shown here were small and did not reach an asymptotic level after 5 h, 20 min of training, which may indicate that subjects were still learning at the point of final testing. This was especially true for the 3OS condition, which proved particularly difficult for the listeners in this study. While the difficulty in learning 3OS appears to contrast with previous studies with monaural basally shifted speech (Rosen *et al.*, 1999), the sparse spectral representation in the processing is a novel feature that has not been previously studied and may have been a source of difficulty. Tyler and Summerfield (1996) showed that cochlear implant patients are still adjusting to their speech processors 6 months and longer after implantation, and the training times considered here were very short by comparison.

It is also possible that adaptation to mismatched frequency-place maps requires experience with both the shifted alone (3OS) and binaurally mismatched (6DOS) maps. For the 6DOS-trained group, training led to larger improvements overall, and there was some evidence of adaptation specific to the shifted bands (3OS). While training with 3OS led to smaller overall improvements, for this group there was evidence of shift-specific learning and better transmission of F2 information in the 6DOS condition. While training with just the shifted components (3OS) alone may be insufficient to allow for a binaural advantage from the mismatched maps, a period of training with just the shifted map may facilitate adaptation to the shifted components, which may only then lead to a binaural advantage with sufficient further training.

Experiment 2 was designed to further explore the effects of type and time course of training on accommodation of mismatched frequency maps. The training period was doubled, and listeners were trained with both the shifted bands alone (3OS) and with these bands in combination with unshifted bands (6DOS) to provide listeners with extensive experience in both conditions.

III. EXPERIMENT 2

A. Method

1. Subjects

Six normally hearing speakers of British English took part and were paid to participate. They all had normal (<20 dB HL) pure-tone thresholds at 0.5, 1, 2, and 4 kHz. None had participated in Experiment 1.

2. Signal processing

Signal processing was the same as for Experiment 1.

3. Training

Training for Experiment 2 was similar to Experiment 1, with the exception that subjects were trained in 30 sessions each of 20 min. Each training session was divided into 10 min blocks which alternated between the 6DOS and 3OS processors. Author K.M., a speaker of Standard Southern British English, was the training talker for Experiment 2.

4. Test materials

Test materials for Experiment 2 included the same IEEE sentence and vowel identification tests from Experiment 1. Because of floor effects with the 3OS condition in Experiment 1, we also used easier high-context BKB (Bench*et al.*, 1979) and IHR (Institute of Hearing Research, MacLeod and Summerfield, 1990) sentences in order to more clearly demonstrate learning with 3OS. These two sentence sets have similar syntactic constructions and are essentially equivalent in intelligibility. The female talker was the same as that for the IEEE sentences, but the male talker was different. The BKB sentences consist of 16 sentences per list with either 3 or 4 keywords per sentence, making 50 keywords per list. The IHR sentences consist of 15 sentences of 3 keywords each, hence 45 keywords per list. BKB sentences were used for familiarization testing, with IHR sentences used for all subsequent testing.

5. Procedure

Table V sets out the testing and training regime used for Experiment 2, which is similar to that of the previous experiment, with the following modifications. A familiarization test session was included at session 1 in which subjects were tested with the IEEE sentences, the Bamford–Kowal–Bench (BKB) sentences, and vowels. This was primarily for familiarization so that any rapid adaptation to the unshifted vocoder processing would occur before training commenced

TABLE V. Sequence of training and testing conditions for Experiment 2.

Session	Training	Processor	Testing
1	5 min AV, 5 min AA	6DU	Familiarization: Unprocessed vowels; *Pre-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), BKB sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 4 conditions × 2 talkers)
2	…	…	*Second Pre-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 4 conditions × 2 talkers)
3	5 min AV × 2, 5 min AA × 2	6DOS+3OS	None
4–17	10 min AA × 2	6DOS+3OS	None
18	…	…	*Mid-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 5 conditions × 2 talkers)
19–33	10 min AA × 2	6DOS+3OS	None
33	…	…	*Post-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 5 conditions × 2 talkers)

(Davis *et al.*, 2005). BKB sentences were used for easy material in only the very first pre-training session. After session 1, only IHR sentences were used for easy sentence material. Also, in the first training session only (session 3 in Table V) the first 5 min with each processor was auditory-visual. All subsequent training for this experiment was auditory only. All other procedural considerations were the same as for Experiment 1.

6. Conditions

Four test conditions were used in Experiment 2: 6DU, 6DOS, 3EU, and 3OS. Because we anticipated ceiling effects with the easy sentence material and the dichotic unshifted (6DU) processor, this condition was not tested with the IHR sentences.

B. Results

1. Test exposure learning

A repeated measures ANOVA was used to compare results between the first two pre-training test sessions. No significant differences between these two test sessions were found for either the IEEE sentence test or the vowel test, which implies that any further improvements beyond the second test session were likely the result of training rather than incidental learning. All subsequent analyses are therefore based on the final three test sessions.

2. IEEE sentences

Scores for the IEEE sentences across training sessions are shown for the two talkers combined in the top left panel of Fig. 8. Performance tended to improve with training in all conditions. An ANOVA on data from the final three test sessions showed significant main effects of number of training sessions [$F(2,10)=29.0$, $p<0.001$] and processor [$F(1,5)=72.2$, $p<0.001$]. While there was a significant session with talker interaction [$F(2,10)=6.85$, $p=0.013$], there was no training session with processor interaction, which indicates that learning through training was similar in all conditions, including 3OS. This is a similar finding to the 3OS-trained group in Experiment 1, who had an equivalent amount of training with the 3OS condition as the subjects in this experiment.

Here there was also direct evidence of learning of the shifted bands (3OS) since a planned contrast of training sessions in this condition showed intelligibility after training was significantly better than before training ($p=0.038$). Yet despite this learning, and the extended period of training in Experiment 2, there was still no evidence of an advantage for the mismatched maps. In *post-hoc* tests, performance with 6DOS and 3EU did not differ, even though there was room for improvement after 10 h of training since 3EU was worse than 6DU ($p=0.02$).

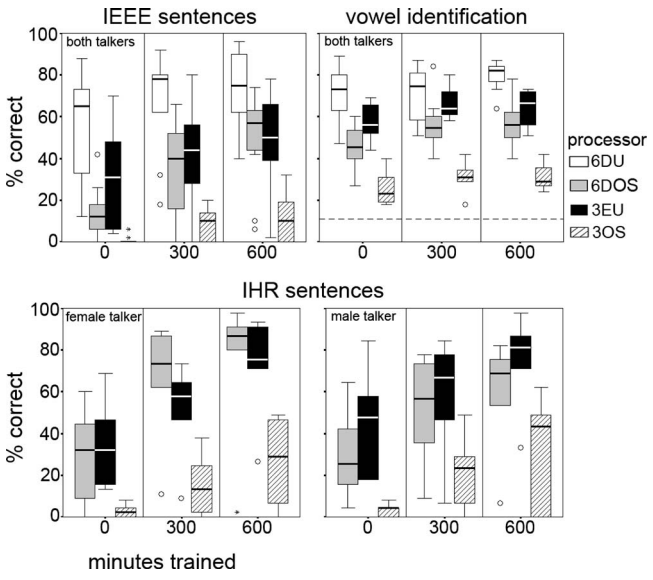


FIG. 8. Performance as a function of amount of training for the final three test sessions of Experiment 2. The top left panel shows IEEE sentence scores, and the top right panel vowel identification. In the bottom two panels are scores for the IHR sentences, which have been split by talker because of a significant talker effect. Training time periods are separated by vertical lines, and for the vowel identification, the horizontal reference line indicates chance performance (11.1%).

3. High-context (IHR) sentences

The bottom panels of Fig. 8 show scores for the IHR sentences for the female and male talkers individually. ANOVA showed a significant interaction between talker and processor [$F(2,8)=6.58$, $p=0.022$], so scores for each talker were analyzed separately. These showed significant effects of training session and condition. The bottom panels of Fig. 8 show that median intelligibility of the 3OS condition improved from close to floor before training to near 40% for both talkers after training, which is a much higher level of intelligibility for this processor than that seen in Experiment 1. A planned contrast showed that improvement with training in this condition was significant ($p=0.02$).

Surprisingly, *post-hoc* tests on the data from all six subjects showed that, for the male talker, post-training intelligibility with 3OS did not differ significantly from that with the 6DOS processor. This may have been the result of increased learning with the 3OS processor. However, performance for the 6DOS condition was significantly worse than the 3EU condition for this talker, which may also indicate interference or distraction from the addition of the shifted channels. Despite clear evidence of adaptation to 3OS, however, subjects did not show an advantage for the mismatched maps since performance with 3EU and 6DOS did not differ significantly after training.

4. Vowel identification

Boxplots of vowel identification for the two talkers combined appear in the top right panel of Fig. 8. ANOVA on the final three test sessions showed significant main effects of session [$F(2,8)=20.4$, $p=0.001$] and processor [$F(3,15)=77.8$, $p<0.001$], but no significant interactions. The lower bound of the 95% confidence interval of the estimated marginal mean was always greater than 11%, indicating that intelligibility was statistically greater than chance for all conditions and at all test points. While a planned contrast showed significant learning of the shifted channels (3OS) with training ($p=0.046$), *post-hoc* testing on the post-training vowel scores revealed no significant difference in intelligibility between the 6DOS and 3EU processors, so subjects did not show a binaural advantage for the mismatched maps, despite having room for improvement to do so (6DU $>$3EU, $p=0.014$). Moreover, a planned contrast showed that post-training vowel intelligibility was marginally worse with the binaurally mismatched processor (6DOS) than with the monaural unshifted bands (3EU). Although this test just missed significance ($p=0.051$), the finding is suggestive of distraction with the addition of the shifted channels, rather than binaural benefit.

As for Experiment 1, the post-training vowel confusion matrices were entered into an information transfer analysis to explore the differences in information transmission with the 6DOS, 3EU, and 3OS processors. The findings are shown in the right panel of Fig. 7. The percent information transmitted was then entered into a repeated measures ANOVA, with within-subjects factors of talker, processor (6DOS, 3EU, and 3OS only), and feature (F1, F2, and duration). The ANOVA revealed no significant effect of talker, but there was a significant interaction of processor with feature [$F(4,20)=4.16$, $p=0.013$]. The F1 information transmitted was significantly lower for 6DOS than for 3EU ($p=0.036$), but there was no difference between these conditions for F2 and duration information. Information transmission with 3OS was always significantly worse than 6DOS and 3EU.

C. Discussion

Subjects in Experiment 2 showed greater signs of adaptation to the shifted portion of the binaurally mismatched processor (3OS) than those in the previous experiment. First, interactions of session with processor were not found for the IEEE sentences and vowel recognition, which indicates that learning through training was similar in all conditions. There was a significant session with processor interaction for the IHR sentences, but this was the result of ceiling effects with 3EU. More significantly, there was direct evidence of learning of 3OS since planned contrasts of sessions within the 3OS condition showed significantly better intelligibility after training with all test materials. Median intelligibility of easy sentences with just the three shifted bands (3OS) after 10 h training improved from near floor to around 40%. In Experiment 1, by contrast, post-training intelligibility with 3OS was near floor with easy sentences for the 3OS-trained group.

Intelligibility with 3OS was only clearly above floor levels with interspersed and/or concurrent training with unshifted vocoded speech: although the 3OS-trained group in Experiment 1 received the same amount of training with that condition as the subjects in Experiment 2, only the latter group of subjects showed such improvements. The greater improvement may have arisen from either more exposure time to the shifted speech (through exposure to both 6DOS and 3OS), from the concurrent and/or interspersed exposure to easier unshifted speech, or a combination of the two. Recently Li *et al.* (2009) showed that unsupervised adaptation to a large basalward shift (8 mm) only occurred with interspersed exposure to a more moderate basalward shift (4 mm). It thus seems plausible that exposure to an easier condition may facilitate the learning of more difficult speech conditions. However, the increased learning of 3OS in Experiment 2 may also have been the result of a change in training talker.

Despite this increased evidence of adaptation to the basally shifted speech and the extended training period allowed for adaptation, again no binaural advantage was found for the binaurally mismatched frequency-place maps. Performance with the 6DOS processor never exceeded that with the three unshifted channels alone (3EU), even though there was clear room for improvement from the addition of the odd channels, since 3EU was typically worse than 6DU. Moreover, there was some evidence of interference from the shifted bands when they were presented together with the unshifted bands since post-training performance with 6DOS was worse than 3EU for the vowels, the male talker IHR sentences, and F1 information transmission. This latter finding is similar to the 3OS-trained group in Experiment 1 and

suggests that explicitly training with the 3OS condition may make the shifted F1 information difficult to ignore when presented with unshifted information in 6DOS.

IV. GENERAL DISCUSSION

A. Resistance to learning

Listeners in both experiments demonstrated evidence of adaptation following a period of training, yet intelligibility with the 6DOS processor never exceeded that with just the three unshifted channels (3EU). The findings thus indicate that listeners are resistant to learning cochlear frequency-to-place maps that differ greatly between the ears. The mechanism involved in adaptation to binaurally mismatched frequency-place maps is likely to be different from that for adaptation to monaural basalward spectral shifts, where significant if partial adaptation is evident after as little as 3 h training (Rosen *et al.*, 1999).

In particular, learning of the 3OS condition proved more difficult than expected on the basis of previous studies. In Experiment 1, improvements in this condition were small. Experiment 2 showed significant improvements, but these only appeared over the longer time course of training and with concurrent and interspersed training with unshifted channels (6DOS), which was somewhat surprising. The smaller number and non-contiguous output of the shifted channels, or differences in the parameters of the vocoding, such as carrier type and smoothing filter, may have influenced the outcome of training. Most simulation studies showing adaptation to upward spectral shifts have employed noise-vocoding (Rosen *et al.*, 1999; Faulkner *et al.*, 2001, 2003; Fu and Galvin III, 2003; Faulkner *et al.*, 2006; Smith and Faulkner, 2006; Stacey and Summerfield, 2007), although there is also evidence of adaptation with sine-vocoding (Li and Fu, 2007; Li *et al.*, 2009). While an earlier study indicated that these should be equivalent in intelligibility (Dorman *et al.*, 1997b), emerging evidence suggests that for small numbers of channels, there are significant differences in the intelligibility of unshifted noise and sine vocoders (Souza and Rosen, 2009), with noise vocoders being more intelligible when envelope smoothing frequency is low (30 Hz, similar to that used here). What may be a comparable resistance to learning of shifted spectral cues has been observed for frequency-lowered amplitude envelope signals when these were, as here, limited to fairly low-rate (50 Hz) modulations and imposed on sinusoidal carriers (Grant *et al.*, 1994). However, that was a study of auditory-visual speech perception so listeners may have had less to gain from learning to reinterpret the frequency-shifted spectral information than would be the case here.

Even if this resistance to learning were in part due to the use of a low smoothing cut-off frequency with sine-vocoding, it can be argued that low-frequency envelope smoothing should be preferred when simulating CI processors because users of current cochlear implant systems are not sensitive to higher-rate modulations. Moreover, the use of noise vocoders here would not have been suitable since overlap of noise carriers between the ears may have led to confounding effects. Notwithstanding these considerations, there was clear evidence of adaptation to the 3OS processor, especially in Experiment 2, yet still no evidence of a binaural advantage, so it seems unlikely that the more difficult 3OS condition is the sole explanation for the lack of binaural advantage.

It is possible that aspects of the training procedure, such as the listener hearing their own voice unprocessed during training, or the listener returning to everyday speech conditions outside of the experiment, may have hindered adaptation. However, the training procedure was identical to that used in previous studies. Rosen *et al.* (1999) showed adaptation to four-channel noise-vocoded speech with a 6.5 mm basalward shift in just 1 h and 20 min, and further studies have shown adaptation in comparably short time periods (Faulkner *et al.*, 2006; Smith and Faulkner, 2006). Though the processing in the present experiments was markedly different from that typically considered in previous adaptation studies (e.g., sparsely sampled output at each ear and binaurally mismatched, sine-excited vocoding), the training times were comparably longer to compensate for the anticipated increased difficulty. Yet despite the extended course of training in these experiments (5–10 h), subjects showed no evidence of a binaural advantage for the binaurally mismatched processing, even though there was evidence of adaptation to the monaural shifted speech, especially in Experiment 2. It thus seems unlikely that the lack of binaural advantage shown in these experiments is entirely attributable to a deficiency in the training procedure.

A novel aspect of the frequency-place mapping examined here, and a possible source of this resistance to learning, is the altered relative frequency order of the output bands in the binaurally mismatched processor. When combinations of shifted and unshifted bands were presented together, as in the 6DOS and 6OS processors, the output consisted of frequency information that had been shifted in both frequency and relative order. Ranking the output with reference to the relative frequency order of the input analysis bands yields, from low to high frequency, (2–1–4–3–6–5). Informative cross-frequency patterns, such as relative frequencies of the first and second formants, may as a result be destroyed. Evidence from the experiments here suggests that, despite the enriched frequency content, listeners do not learn to combine this altered output in the same way as for matched frequency-place maps. If the shifted channels contributed to intelligibility with 6DOS at all, they appeared to serve as a distraction, with the transmission of first formant information being particularly vulnerable.

The lack of binaural advantage for combinations of shifted and unshifted bands shown here may be indicative of a constraint on plasticity for speech perception to cases where relative frequency order has been preserved. This would be consistent with the finding that a frequency-warped but order-preserving frequency-to-place map is relatively easy to learn (Smith and Faulkner, 2006). There is, however, some evidence to the contrary: in his seminal study on adaptation to spectrally altered speech, Blesser (1972) trained listeners with speech that had been spectrally rotated around 1600 Hz so that low frequencies became high and vice versa. Even under such drastic conditions, some listeners did even-

tually learn to converse, although this did not appear until much later in the study, which consisted of 20 45 min sessions. Azadpour (2008) also found that listeners could adapt to aspects of spectrally rotated speech over a similar time course to that in Experiment 1.

With spectrally rotated speech, however, the relative ordering of frequency information is intact, even if it has been inverted. By contrast, the processing used here may selectively invert speech cues such as vowel formant patterns which may result in a novel percept, in addition to cues being presented to the wrong frequency region. In the present investigations, this appeared to hinder adaptation. Specifically, in Experiment 1, there was evidence of shift-specific adaptation to the 3OS condition for both groups for the vowel test, yet listeners still did not show a binaural advantage when listening in the 6DOS condition. This was despite the fact that in the absence of shifting, the three even (3EU) and three odd (3OU) processors resulted in significantly different vowel confusions, which would imply that there was room for improvement by combining the information if the listener had been able to do so.

B. Implications for bilateral stimulation with cochlear implants

Since performance did not reach asymptote in either of the present experiments, it remains to be determined whether continued learning beyond that demonstrated here would be shown, or whether this is too difficult a mapping to learn fully. In this respect, it is difficult to interpret the implications of this resistance for cochlear implant patients because the extent of their experience with any clinically fitted processor will be far greater than the 5–10 h examined in the laboratory here. For example, the group of patients investigated by Tyler and Summerfield (1997) reached asymptotic performance an average of 30–40 month post implantation.

Several studies have shown that some patients experience a synergistic improvement when using both implants over either implant on its own, even for speech in a quiet laboratory (Dorman and Dahlstrom, 2004; Litovsky *et al.*, 2006). The drastic spectral mismatch examined here may arguably be more representative of the mismatch with acoustic hearing and may be uncommon between two bilateral electrode arrays of the same design. Smaller mismatches than that studied here may be tolerated by CI users. Yet, not all bilateral CI users show an advantage over their best ear alone (Wackym *et al.*, 2007; Tyler *et al.*, 2007) and substantially mismatched frequency-place maps underpin this lack of advantage.[1] In a recent study of patients with bilateral cochlear implants, Mosnier *et al.* (2009) found that bilateral CI patients with asymmetrical monaural CI speech scores did not show a binaural advantage for speech in quiet, which suggests a difficulty with mismatched frequency-place maps.

Since the processor design considered here did not directly simulate typical bilateral implant systems currently in use, there may be limitations to generalizing the findings to CI patients. The speech processors were designed to minimize redundancy between the two ears in order to investigate global aspects of binaural place mismatch in the absence of local conflicts due to (1) the presentation of the same information to two different places or to (2) the presentation of different information to the same tonotopic location in each ear. However, typical bilateral cochlear implant systems do deliver roughly the same signal to each ear. Interleaving channels between the ears may have left insufficient cues to bind unshifted and shifted channels into a single auditory object. An increase in signal redundancy provided with "typical" bilateral processing may facilitate auditory grouping and thus the binaural combination of mismatched maps. However, the more spectrally rich signal provided by the better ear (i.e., the ear with the smallest degree of shift) with non-interleaved processing could also lead listeners to learn to attend to the better ear alone. Further research could explore whether amplitude comodulation of the shifted and unshifted channels would facilitate grouping of the mismatched signals as a single auditory object and thus allow for a binaural advantage (Carrell and Opie, 1992).

Current bilateral cochlear implants are fitted by mapping the full spectrum of speech information to each ear, but this is not a requirement for robust speech perception, as is clearly demonstrated by the speech intelligibility advantage with electro-acoustic stimulation (Ching, 2005). Alternative frequency-place mapping strategies could seek to align the frequency range delivered to each ear through assigning processor filter frequencies that were matched according to interaural location. For cases of asymmetrical insertion depths, this would likely entail that the highest and lowest channels were presented to only one ear. Such an alternative processing strategy would avoid drastic mismatches of frequency-place maps between the ears while maximizing signal redundancy and may ultimately lead to a stronger binaural advantage. To this end, robust methods for matching frequency-place maps between the ears for humans would need to be developed. New research into more objective techniques, such as the binaural interaction component of the evoked auditory brainstem response (EABR), as described by Smith and Delgutte (2007), could prove promising.

Many of the benefits of binaural hearing are realized for speech in noise, where differences in the sound signal at each ear can be used to obtain a better representation of what has been said. Indeed, of the two subjects tested by Dorman and Dahlstrom (2004), one showed binaural benefit over the better ear only in the noise condition. Better-ear effects may dominate in the absence of noise and thus limit any potential binaural advantage from the mismatched signals at each ear. If the signal here were presented in noise, listeners may have stronger motivation to rely on the shifted channels. When testing in noise, it may also be possible to use more spectral bands without achieving ceiling effects with the unshifted speech, and this increased spectral resolution may, in turn, facilitate learning.[2]

Current cochlear implant systems do not contribute the fine structure information thought to underpin many binaural hearing phenomena in normal hearing listeners. However, Long and colleagues showed at least one bilateral cochlear implantee to be sensitive to the interaural time difference (ITD) and interaural level difference (ILD) cues provided solely through envelope information. Crucially, this was limited to cases where interaural electrode pairs were matched

in frequency (Long *et al.*, 2003). Normal hearing listeners also demonstrate best ITD (Nuetzel and Hafter, 1981) and ILD (Francart and Wouters, 2007) detection for similar interaural cochlear place regions. Though perhaps through a different underlying mechanism than has been considered here, evidence from psychophysics is thus consistent with the present research and suggests that maintaining similar frequency-place maps between the ears may be important for optimizing bilateral CI fittings.

V. CONCLUSION

This research examined whether listeners could learn to demonstrate a binaural advantage for speech information with a unilateral spectral shift. Despite undergoing significantly longer training periods than in previous studies with monaural spectral shifts, subjects in the present investigations never showed a binaural advantage for the binaurally mismatched frequency-place map. Post-training performance with the binaurally mismatched processor never exceeded that with the three unshifted channels, even after 10 h of training. This resistance to learning is suggestive of a constraint on speech perceptual plasticity to instances where relative frequency order is preserved. It may thus be important to keep frequency-place maps similar in the two ears when optimizing bilateral CIs for speech perception. The results are consistent with psychophysical studies in bilateral cochlear implant patients, which have also called for matching interaural electrode pairs in order to restore ITD and ILD sensitivities.

ACKNOWLEDGMENTS

Work supported by a grant from the European Commission (Grant No. FP6-004171 HEARCOM), Study Assistance from the Staff Development and Training Unit, Department of Human Resources, UCL awarded to C.M.S., and a Wellcome Trust Summer Vacation Scholarship to K.M. (Grant No. VS/06/UCL/A12).

[1]Evidence from studies with a self-selecting tool for frequency-place mappings supports this conclusion. Fitzgerald *et al.* (2007) found that normal hearing subjects, when listening to simulations of CI processing, preferred frequency-place maps that minimized the mismatch between the ears, regardless of spectral shift.

[2]In a related study, Long *et al.* (2004) used 12 channel binaurally interleaved vocoders similar to those described here, with one ear processed in noise. Listeners in their experiment demonstrated a binaural advantage for consonant perception over either ear on its own. This could not be attributed to binaural redundancy since the interleaved channels ensured that each ear was getting different information. However, the processor in their study was tonotopically matched in both ears. It merits exploration whether this advantage remains if a spectral shift is introduced to one ear.

Azadpour, M. (**2008**). "Perceptual learning in speech processing: Comprehending spectrally-rotated speech," Ph.D. dissertation, University of Trieste, Trieste, Italy.

Bench, J., Kowal, A., and Bamford, J. M. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children," Br. J. Audiol. **13**, 108–112.

Blesser, B. (**1972**). "Speech perception under conditions of spectral transformation I. Phonetic characteristics," J. Speech Hear. Res. **15**, 5–41.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

Carrell, T., and Opie, J. M. (**1992**). "The effect of amplitude comodulation on auditory object formation in sentence perception," Percept. Psychophys. **52**, 437–445.

Ching, T. Y. C. (**2005**). "The evidence calls for making binaural-bimodal fittings routine," Hear. J. **58**, 32–41.

Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Ching, T. Y. C., Incerti, P., Hill, M., and van Wanrooy, E. (**2006**). "An overview of binaural advantages for children and adults who use binaural/bimodal hearing devices," Audiology and Neurotology **11**, 6–11.

Cutting, J. E. (**1976**). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psychol. Rev. **83**, 114–140.

Davis, M. H., Johnsrude, I. S., and Hervais-Adelman, A. T. K. M. C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

De Filippo, C. L., and Scott, B. L. (**1978**). "A method for training and evaluating the reception of ongoing speech," J. Acoust. Soc. Am. **63**, 1186–1192.

Dorman, M. F., and Dahlstrom, L. (**2004**). "Speech understanding by cochlear-implant patients with different left- and right-ear electrode arrays," Ear Hear. **25**, 191–194.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). "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.

Faulkner, A. (**2006**). "Adaptation to distorted frequency-to-place maps: Implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation," Audiology and Neurotology **11**, 21–26.

Faulkner, A., Rosen, S., and Norman, C. (**2006**). "The right information may matter more than frequency-place alignment: Simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion," Ear Hear. **27**, 139–152.

Faulkner, A., Rosen, S., and Stanton, D. (**2003**). "Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth," J. Acoust. Soc. Am. **113**, 1073–1080.

Faulkner, A., Rosen, S., and Wilkinson, L. (**2001**). "Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant processor," Ear Hear. **22**, 431–438.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech Lang. Hear. Res. **40**, 1201–1215.

Fitzgerald, M. B., Tan, C.-T., and Svirsky, M. A. (**2007**). "Customized selection of frequency maps in an acoustic simulation of a cochlear implant, in Conference on Implantable Auditory Prostheses, Tahoe City, CA.

Francart, T., and Wouters, J. (**2007**). "Perception of across-frequency interaural level differences," J. Acoust. Soc. Am. **122**, 2826–2831.

Fu, Q.-J., and Galvin, J. J. III (**2003**). "The effects of short-term training for spectrally mismatched noise-band speech," J. Acoust. Soc. Am. **113**, 1065–1072.

Fu, Q.-J., Galvin, J. J. III, Wang, X., and Nogaki, G. (**2005a**). "Moderate auditory training can improve speech performance of adult cochlear implant patients," ARLO **6**, 106–111.

Fu, Q.-J., Nogaki, G., and Galvin, J. J. III (**2005b**). "Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation," J. Assoc. Res. Otolaryngol. **6**, 180–189.

Fu, Q.-J., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. III (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Grant, K. W., Braida, L. D., and Renn, R. J. (**1994**). "Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech," J. Acoust. Soc. Am. **95**, 1065–1073.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species — 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hamzavi, J., Pok, S. M., Gstoettner, W., and Baumgartner, W.-D. (**2004**). "Speech perception with a cochlear implant used in conjunction with a

hearing aid in the opposite ear," Int. J. Audiol. **43**, 61–65.
Iwaki, T., Matsushiro, N., Mah, S.-R., Sato, T., Yasuoka, E., Yamamoto, K.-I., and Kubo, T. (**2004**). "Comparison of speech perception between monaural and binaural hearing in cochlear implant patients," Acta Oto-Laryngol. **124**, 358–362.
Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). "In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays," Ann. Otol. Rhinol. Laryngol. **107**, 1–16.
Lawson, D., Brill, S., Wolford, R., Wilson, B., and Schatzer, R. (**2000**). Speech Processors for Auditory Prostheses: Ninth quarterly progress report.
Li, T., and Fu, Q.-J. (**2007**). "Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels," J. Assoc. Res. Otolaryngol. **8**, 32–41.
Li, T., Galvin, J. J. III, and Fu, Q.-J. (**2009**). "Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech," Ear Hear. **30**, 238–249.
Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (**2006**). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," Ear Hear. **27**, 714–731.
Loizou, P. C., Mani, A., and Tu, Z. (**2003**). "Dichotic speech recognition in noise using reduced spectral cues," J. Acoust. Soc. Am. **114**, 475–483.
Loizou, P. C., Dorman, M. F., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.
Long, C. J., Deeks, J. M., Leow, M.-C., and Carlyon, R. P. (**2004**). "Simulations of bilateral cochlear implant speech reception in noise," in Association for Research in Otolaryngology Midwinter Meeting, Daytona Beach, FL.
Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.
Macleod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation and recommendations for use," Br. J. Audiol. **24**, 29–43.
Mani, A., Loizou, P. C., Shoup, A., Roland, P., and Kruger, P. (**2004**). "Dichotic speech recognition by bilateral cochlear implant users," International Congress Series, **1273**, 466–469.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Mills, J. H., Dubno, J. R., and He, N. J. (**1996**). "Masking by ipsilateral and contralateral maskers," J. Acoust. Soc. Am. **100**, 3336–3344.
Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., Ambert-Dahan, E., and Ferrary, E. (**2009**). "Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients," Audiology and Neurotology **14**, 106–114.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Robinson, D., and Dadson, R. S. (**1956**). "A redetermination of the equal-loudness relations for pure tones," Br. J. Appl. Phys. **7**, 166–181.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.
Rothauser, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.
Rubin, B. A., Uchanski, R. M., and Braida, L. D. (**1992**). "Integration of acoustic cues for consonants across frequency bands (A)," J. Acoust. Soc. Am. **91**, 2473–2474.
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., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (**2002**). "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients," J. Assoc. Res. Otolaryngol. **3**, 332–350.
Smith, M. W., and Faulkner, A. (**2006**). "Perceptual adaptation by normally hearing listeners to a simulated "hole" in hearing," J. Acoust. Soc. Am. **120**, 4019–4030.
Smith, Z. M., and Delgutte, B. (**2007**). "Using evoked potentials to match interaural electrode pairs with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **8**, 134–151.
Souza, P., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.
Stacey, P., and Summerfield, Q. (**2007**). "Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech," J. Acoust. Soc. Am. **121**, 2923–2935.
Stakhovskaya, O., Sridhar, D., Bonham, B., and Leake, P. (**2007**). "Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants," J. Assoc. Res. Otolaryngol. **8**, 220–233.
Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., Helms, J., and Suárez, H. (**2004**). "Long-term auditory adaptation to a modified peripheral frequency map," Acta Oto-Laryngol. **124**, 381–386.
Tyler, R. S., and Summerfield, Q. (**1996**). "Cochlear implantation: Relationships with research on auditory deprivation and acclimatization," Ear. Hear. **17**, 38S–50S.
Tyler, R. S., Dunn, C. C., Witt, S. A., and Noble, W. G. (**2007**). "Speech perception and localization with adults with bilateral sequential cochlear implants," Ear Hear. **28**, 86S–90S.
van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.
Wackym, P. A., Runge-Samuelson, C. L., Firszt, J. B., Alkaf, F. M., and Burg, L. S. (**2007**). "More challenging speech-perception tasks demonstrate binaural benefit in bilateral cochlear implant users," Ear Hear. **28**, 80S–85S.

Two-microphone separation of speech mixtures based on interclass variance maximization

Maximo Cobos[a] and Jose J. Lopez
Institute of Telecommunications and Multimedia Applications, Universidad Politécnica de Valencia, Valencia 46022, Spain

(Received 28 May 2009; revised 11 November 2009; accepted 24 December 2009)

Sparse methods for speech separation have become a discussed issue in acoustic signal processing. These sparse methods provide a powerful approach to the separation of several signals in the underdetermined case, i.e., when there are more sources than sensors. In this paper, a two-microphone separation method is presented. The proposed algorithm is based on grouping time-frequency points with similar direction-of-arrival (DOA) using a multi-level thresholding approach. The thresholds are calculated via the maximization of the interclass variance between DOA estimates and allow to identify angular sections, wherein the speakers are located with a strong likelihood. These sections define a set of time-frequency masks that are able to separate several sound sources in realistic scenarios and with little computational cost. Several experiments carried out under different mixing situations are discussed, showing the validity of the proposed approach. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3294713]

PACS number(s): 43.72.Ne [WMC] Pages: 1661–1672

I. INTRODUCTION

Humans are surrounded by sound. If we try to concentrate and listen carefully to the things that happen in our environment, probably we will be able to identify more than one source of sound. This fact reveals an important property of human hearing: the ability to distinguish individual sound sources from complex mixtures of sound. When talking about the perception of speech in complex acoustic environments, this human ability is usually related to the well known *cocktail party problem*, first described by Cherry in 1953.[1] The cocktail party effect describes the ability to focus one's listening attention on a single talker among a mixture of conversations and background noises, ignoring other conversations and enabling us to talk in a noisy place.[2] The complex understanding of sound scenes is so familiar to humans that we usually take it for granted and we perform this task without being aware of it. However, computational methods aimed at imitating our hearing sense and brain for understanding complex sound material are far from performing with such an accuracy and are still an active research line that involves mathematical, physical, and psychological issues. In this context, a common approach to achieve understanding of complex sound scenes is usually sound source separation. The solution to the problem of sound separation in real environments is essential for many applications, including automatic speech recognition, music information retrieval, and hearing-aid systems.[3]

The task of estimating and recovering independent source signals from a set of mixtures obtained by one or several observation channels is known as *blind source separation*. Most of the existing separation algorithms are based on statistical assumptions, mainly, sources statistically independent and non-Gaussian. These assumptions often lead to the *independent component analysis* approach.[4] These algorithms have shown to be successful in the linear complete case, when as many observations as sources are available. When the mixing process is convolutive, each sensor observes the original source signals convolved with the impulse response between each source and sensor, making the estimation of the sources even more difficult.[5] This is the case of real mixtures recorded in a room by a set of microphones. Methods based on independent component analysis for convolutive mixtures often apply the separation algorithm separately in each frequency bin using a time-frequency transformation of the observed mixtures. This approach introduces the well-known permutation problem: the different frequency components of the signals are "swapped" and require an alignment process.[6] Moreover, when there are more sources than observation channels, the problem is *underdetermined* (or degenerate), and other properties of the sources such as *sparsity* are exploited. In a strict sense, sparsity means that most signal components are zero. In a practical sense, sparsity means that most signal components are relatively small.[7] Sparsity and overcomplete dictionaries have been discussed in the literature with the aim of giving a solution to the underdetermined problem.[8] Lewicki and Sejnowski[9] proposed an algorithm for learning overcomplete representations of the sources via *maximum a posteriori* estimation. Sparse recovery of the sources was also addressed via l_1-norm minimization by Donoho.[10] When dealing with speech mixtures, it has been shown that they are sparser in the time-frequency domain than in the time domain.[11] Yilmaz *et al.*[12] assumed that the sources are disjoint in the time-frequency domain, i.e., there exists only one source in a given time-frequency point. This assumption leads to the *time-frequency masking* approach. Algorithms based on time-frequency masking have shown to provide significant results in the separation of realistic mixtures of speech; however,

[a]Author to whom correspondence should be addressed. Electronic mail: mcobos@iteam.upv.es

most approaches suffer from the dependency on initial cluster centroids[13] or the accuracy of a peak detection step.[12]

In this paper, we present a source separation algorithm based on time-frequency masking. Inspired by image segmentation techniques, the separation is achieved by using a *maximum interclass variance*[14] criterion between the angular distribution of the sources. Maximization of interclass variance was recently shown to provide successful results in the separation of instantaneous music mixtures.[15] We demonstrate how under the same criterion, it is possible to achieve fast separation of simultaneous speakers in reverberant scenarios by identifying different angular areas wherein the speakers are located with a strong likelihood. These areas are defined by a set of thresholds that, applied to the distribution of DOA estimates in the time-frequency domain, provide the binary masks that separate the sources. We present a complete set of examples using real speech mixtures, considering different degrees of reverberation, different number of sources, and a range of angular separations of speakers. The results show the capability of our approach to separate different speech sources in real situations with minimal computational cost. In addition, the method does not need to perform a previous estimation of the mixing matrix nor to search for peaks or specify initial values, which results in a more robust reverberation and unsupervised algorithm.

This paper is structured as follows: in Sec. II, the description of the source separation problem is presented. In Sec. III, the proposed approach is described in detail. In Sec. IV, some experiments with speech mixtures are presented, and a performance evaluation is carried out considering different mixing situations. Finally, in Sec. V, the conclusions of this work are summarized.

II. THE SOURCE SEPARATION PROBLEM

In the last years, increasingly attention has been given to the source separation problem.[16] The term *blind* is usually employed when very little information about the sources and the mixing process is known. In some situations, recovering the sources from the recorded mixtures or segregating a particular source is very desirable. This can also reveal information about the mixing system that may be useful for other applications. The simplest mixing model is the instantaneous model, where the different mixture channels are formed by linear combinations of the sources. However, in real applications, the mixing process is more complex: each mixture is the sum of all the original source signals filtered by their corresponding source-to-sensor acoustic channel. The impulse response of this channel carries the information of the acoustic environment where the signals have been mixed.

In this section, we review the usual mixing model used for the separation of realistic mixtures of speech and describe the time-frequency masking approach that is effective for separating sparse sources.

A. Mixing model

In this work we consider a two-microphone array setup. The signals recorded by a pair of microphones, $i=1,2$, in a room where N sound sources are present can be modeled as a finite impulse response convolutive mixture, written as

$$x_i(t) = \sum_{j=1}^{N} \sum_{\ell=0}^{L_m-1} h_{ij}(\ell) s_j(t-\ell), \quad i=1,2, \tag{1}$$

where $x_i(t)$ is the signal recorded at the ith microphone at time sample t, $s_j(t)$ is the jth source signal, $h_{ij}(t)$ is the impulse response of the acoustic path from source j to sensor i, and L_m is the maximum length of all impulse responses. More generally, the mixing process can be written as

$$x_i(t) = \sum_{j=1}^{N} s_{ij}(t), \quad i=1,2, \tag{2}$$

where s_{ij} is called the image of the jth source in the ith microphone

$$s_{ij}(t) = \sum_{\ell=0}^{L_m-1} h_{ij}(\ell) s_j(t-\ell). \tag{3}$$

A simplification of the convolutive model can be found in the particular case when only the direct path from each source to each sensor is considered. In this case, only an arrival delay τ_{ij} between each source-sensor pair is sufficient to explain the generation of the mixtures. Thus, the mixing filters of the model become $h_{ij}(t)=a_{ij}\delta(t-\tau_{ij})$, resulting in the anechoic mixing model

$$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t-\tau_{ij}), \quad i=1,2,\ldots, \tag{4}$$

where τ_{ij} and a_{ij} are the arrival delay and attenuation factor between source s_j and sensor i, respectively.

It is here useful for Secs. III and IV to rewrite the above model in the time-frequency domain, considering the short-time Fourier transform (STFT). This transform divides a time-domain signal into a series of small overlapping pieces; each of these pieces is windowed and then individually Fourier transformed.[17] Due to the linearity of the STFT, the above model can be expressed as

$$X_i(k,m) = \sum_{j=1}^{N} a_{ij} S_j(k,m) e^{-j2\pi f_k \tau_{ij}}, \quad i=1,2, \tag{5}$$

where k and m are the frequency and time frame indices, respectively, $S_j(k,m)$ denotes the STFT of the source signal $s_j(t)$, f_k is the frequency corresponding to frequency index k, and τ_{ij} is the time delay corresponding to the path from source s_j to sensor i. Throughout this paper, the above simplified model will be considered.

B. Time-frequency masking

Signal decompositions or additive expansions have been shown to provide a sparse representation of audio signals, which is a desirable feature for many speech processing systems. A signal is said to be sparse if most of its components are zero or close to zero. In the source separation context, the advantage of a sparse signal representation is clear: the sparser a signal is, the less probability it will overlap with

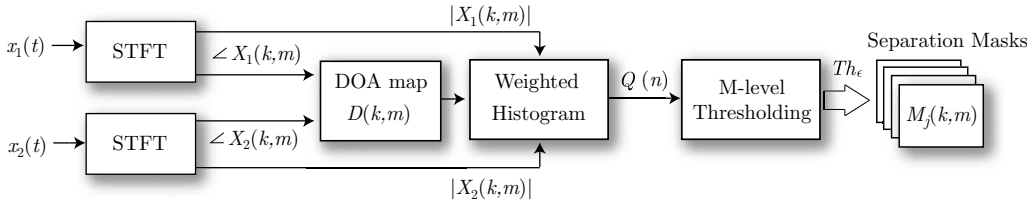


FIG. 1. Block diagram of the separation method.

other signals in a mixture.[18] Time-frequency representations have been widely used with the aim of working with sparse sources, being the STFT the most common transform used for the separation of underdetermined speech mixtures.

The problem of separating more sources than mixtures is not only related to the difficulty of estimating more mixing parameters from a reduced set of data, but in providing a meaningful solution to the system of Eq. (4). In fact, the sparsity provided by time-frequency representations is sufficient to estimate the complete set of mixing parameters with relatively good accuracy,[12] but some criterion is still required to recover the sources from the infinite possible solutions to the problem.

In this context, one of the most simple and powerful approaches for recovering multiple source signals from a mixture is time-frequency masking. It should be noted that the term "masking" here means weighting (filtering) the mixture, which is different from the same term used in psychoacoustics, where it means blocking the target sound by using acoustic interference. If the sources are assumed to be W-disjoint orthogonal (WDO), it is likely that every time-frequency point in the mixture with significant energy is dominated by the contribution of only one source. The "W" in WDO refers to the analysis window used in the time-frequency transform.[19] This can be mathematically expressed as

$$S_j(k,m)S_{j'}(k,m) = 0, \quad \forall\, j \neq j', \quad \forall\, k,m. \tag{6}$$

The above equation means that two different sources never share energy in the same time-frequency point. Therefore, time-frequency masking approaches attempt to identify the dominating source in each time-frequency bin (k,m), obtaining a binary mask (filter) $M_j(k,m)$ that separates the jth source from the mixture. The estimates of the spatial images of the sources can be directly obtained by applying the masks to the input mixtures

$$\hat{S}_{ij}(k,m) = M_j(k,m)X_i(k,m), \quad \forall\, k,m. \tag{7}$$

The estimates in the time domain $\hat{s}_{ij}(t)$ are obtained applying the inverse STFT operator. Further processing methods can be used for improving the isolation of the sources by means of refined time-frequency masks.[20]

1. The ideal binary mask

Consider the sum of all the signals that interfere with source s_j in the STFT domain

$$U_j(k,m) = \sum_{j'=1, j' \neq j}^{N} S_{j'}(k,m), \quad \forall\, k,m. \tag{8}$$

The *ideal binary mask* for a source $S_j(k,m)$ is defined as the binary time-frequency mask that is 1 for bins where its energy is higher than all the interfering sources:

$$M^{\text{ideal}}(k,m) = \begin{cases} 1 & \text{if } 20\log\left(\dfrac{|S_j(k,m)|}{|U_j(k,m)|}\right) \geq 0 \\ 0 & \text{elsewhere} \end{cases} \quad \forall\, k,m. \tag{9}$$

This mask has been shown to be optimal when applied to a sound mixture,[12] and that is why the ideal binary mask has been suggested as a major computational goal of sound source separation algorithms. This is specially true in *computational auditory scene analysis*,[21] since it has proven to be highly effective for robust automatic speech recognition systems and improved human speech intelligibility in noise.[22]

III. SEPARATION FRAMEWORK

An overview of the different stages of the separation algorithm is depicted in Fig. 1. The approach presented in this paper can be summarized in the following steps:

1. *DOA map calculation.* The input channels are transformed into the STFT domain. The phase information in each time-frequency point is used for estimating the DOA of a given frequency in the current analysis window. A DOA map matrix is formed with the estimations obtained from the analyzed signal.
2. *Coherence test and weighted histogram.* A coherence test is performed with the aim of identifying reliable DOA estimates and increasing the robustness against reverberation. In addition, robustness is further improved by using the STFT magnitude of both input channels to construct an amplitude-weighted histogram. This histogram is the input for the multi-level thresholding algorithm.
3. *Multi-level thresholding.* The histogram obtained in the previous step is processed for calculating a set of thresholds that maximizes the interclass variance of the time-frequency points according to their estimated DOA. These thresholds are used for segmenting the DOA map into the final separation masks.

In Secs. IIIA–IIIC the above steps are described in detail.

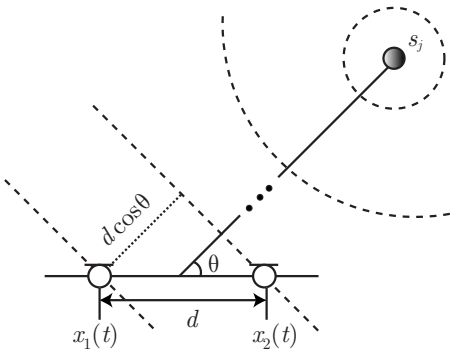


FIG. 2. Two microphones capturing a signal from a source s_j with direction-of-arrival θ. The path difference between both sensors is $d\cos\theta$.

A. DOA map

The DOA map $D(k,m)$ is calculated from the phase difference in the two microphone signals. Figure 2 shows a pair of microphones, capturing the signal arriving from a source s_j with angle θ. The inter-microphone distance is d. Assuming that the source is sufficiently distant to consider plane wavefront incidence, the path difference between the two microphones is $d\cos\theta$.

Given the model of Eq. (5), the relation between the phase difference of the two sensors and the incidence angle of the source for frequency f_k and time frame m is given by the next equation:

$$\angle X_2(k,m) - \angle X_1(k,m) = 2\pi f_k(\tau_{2j} - \tau_{1j}) = 2\pi f_k\left(\frac{d\cos\theta}{c}\right), \tag{10}$$

where $\angle(\,)$ denotes the phase of a complex number and c is the speed of sound ($\approx$340 m/s).

The DOA map represents the cosine of the DOA ($\cos\theta$) for each frequency f_k at each time window m, and it is calculated from the STFT of the mixture channels as

$$D(k,m) = \frac{c}{2\pi f_k d}\angle\left(\frac{X_1(k,m)}{X_2(k,m)}\right), \quad \forall\, k,m. \tag{11}$$

Note that frequencies above $f_k > c/2d$ will suffer from spatial aliasing.[12] Nevertheless, speech sources concentrate most of their energy below 5 kHz and small distances ($d \sim 3$ cm) easily satisfy this condition for the effective bandwidth of speech.

In Figs. 3(a) and 3(b) the left and right channel spectrograms of a mixture of three speakers in an anechoic environment are represented. In this mixture, the speakers are uniformly distributed in the azimuth plane. The amplitude relationship (in dB) between the two channels is represented in Fig. 3(c). Note that very little information is provided by amplitude differences when the microphones are very close to each other. In fact, when the inter-microphone distance verifies $d \ll D$, being D the distance between the source and the array, then $a_{1j} \approx a_{2j}$, and so the modulus $|X_1(k,m)|/$

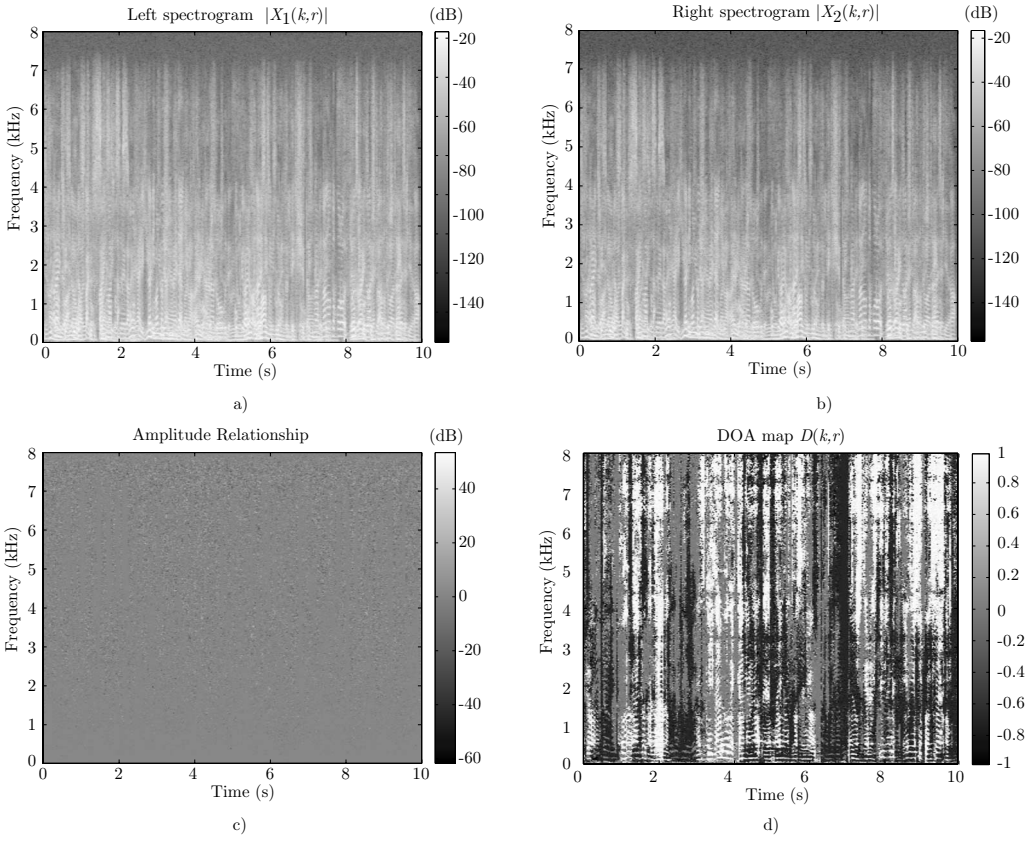


FIG. 3. (a) Left observation channel spectrogram; (b) right observation channel spectrogram; (c) amplitude relationship between the two channels, calculated as $20\log_{10}(|X_1(k,m)|/|X_2(k,m)|)$; and (d) DOA map of the mixture.

$|X_2(k,m)| \approx 1$. However, the DOA map extracted from the phase information of the mixtures [Fig. 3(d)] shows clear zones related to the activity of the three speakers in the time-frequency domain. It is also important to notice that the use of the phase difference as localization cue has another advantage over the amplitude ratio, since it avoids the division by zero when $|X_2(k,m)|=0$.

Under anechoic conditions, the different zones corresponding to each speaker are highly visible as different gray-level zones, which support the presence of predominant speech signals. However, when the mixing is convolutive, multiple reflections of the source signal arrive at both microphones. These reflections cause deviations in the DOA estimation and the zones corresponding to each speaker are harder to identify. In Sec. IV, the performance of the algorithm will be evaluated considering different degrees of reverberation with the aim of showing the degradation introduced by room reflections.

B. Coherence test and weighted histogram

1. Interaural coherence and coherence tests

If a single source in free-field is considered, sound from only one direction arrives at the microphones with a deterministic *interaural time difference*, and thus, the DOA of the source can be easily determined by the observed phase difference, as explained in Sec. III A. However, in complex listening situations, i.e., in the presence of other sound sources or room reflections, it often occurs that sound from different directions concurrently reaches the sensor array, resulting in erroneous DOA estimations.

The robustness of the method against these complex environments can be improved by discarding time-frequency bins where other sources and reverberation are dominant. This is done by means of a coherence test. In fact, from an auditory perspective, *interaural coherence* plays an important role on the localization of sound events in complex listening situations. Thus, the degree of "similarity" between the left and right ear entrance signals has been shown to provide a very important cue for dealing with the localization of sources in difficult environments. A measure for interaural coherence usually has a range between 0 and 1, where 0 means that two signals are coherent (signals are equal with possibly a different scaling and delay) and 1 means that the signals are independent.[23] For example, Faller proposed a successful localization system based on interaural time differences and coherence estimated from the normalized cross-correlation function.[24] From an algebraic perspective, Mohan *et al.*[25] proposed recently a coherence test in the time-frequency domain based on the identification of low-rank bins, where the time-frequency covariance matrix of the input mixtures has effective low-rank (ideally 1 when a time-frequency point is dominated by only one source). Alternatively, Avendano proposed the use of the following *short-time coherence function*[26]

$$\Phi(k,m) = \frac{|\Phi_{12}(k,m)|}{\sqrt{\Phi_{11}(k,m)\Phi_{22}(k,m)}}, \tag{12}$$

where the statistics $\Phi_{ii'}(k,m)$ are a practical way of computing the inter-channel correlation $E\{X_i(k,m)X^*_{i'},(k,m)\}$, given by

$$\Phi_{ii'}(k,m) = (1-\lambda)\Phi_{ii'}(k,m-1) + \lambda X_i(k,m)X^*_{i'}(k,m), \tag{13}$$

where * denotes complex conjugation. The classical coherence function is defined for wide-sense stationary random processes.[27] However, due to the non-stationarity of speech, this function introduces a forgetting factor $\lambda \in [0,1]$ to compute for the cross-correlation between the observation channels on a block of time frames, which can be adjusted to balance between current and previous estimations. The coherence function $\Phi(k,m)$ has values close to 1 in time-frequency regions where a source is present, and it is usually smaller when sounds from different directions overlap. Therefore, high-coherence bins above a certain threshold Φ_T are selected as having reliable DOA estimates (for example, $\Phi_T > 0.9$). This pre-selection is used to identify time-frequency points that are more likely to give a better estimation of the DOA, which results in a sharper histogram, as it will be seen in Sec. III B 2.

2. Amplitude-weighted histogram

The use of histograms for studying phasic behaviors of microphone arrays has already been discussed in the literature, for example, in the use of *delayograms*.[28] Histograms provide a simple way for roughly estimating an underlying probability density function based on observed data. In our approach, the histogram obtained from the DOA estimates in the time-frequency domain reveals important information about the angular distribution of the sources and makes it able to carry out the separation, considering this distribution as a mixture of classes that are angularly spaced.

As commented in Sec. III B 1, the effect of a non-ideal scenario affects negatively the estimation of the DOA at each time-frequency bin. This is due to the phase distortion caused in the observation channels as a result of other source contributions and reverberation. If the phase is not very much altered in a given time-frequency bin, the estimated DOA calculated from the phase difference at that bin will be similar to the real one. In contrast, if the phase has been distorted by other sources or room reflections, the quality of the estimation will be poorer. This effect is depicted in Fig. 4. In this figure, it can be seen the addition of two phasors. One phasor represents the source of interest s_d and the other, s_i, represents the total contribution of the interference (other sources and reflections). It can be observed that the phase deviation $\Delta\phi$ caused by the interference is highly related to the magnitude of the observation channel. In Fig. 4(a), the deviation produced by the total interference is significantly smaller than in Fig. 4(b).

Considering the above relationship between magnitude and phase stability in a given time-frequency bin, a more accurate segmentation of the DOA map will be achieved if

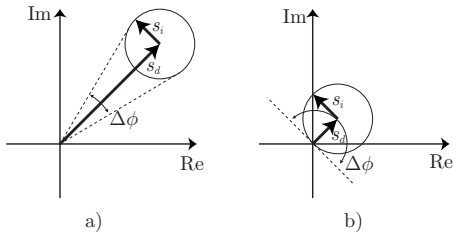


FIG. 4. (a) Possible phase deviation $\Delta\phi$ when the source of interest s_d has big amplitude in comparison to the total interference s_i; (b) possible phase deviation $\Delta\phi$ when the source of interest s_d has low amplitude in comparison to the total interference s_i.

points with high magnitude are given a greater weight in the formation of the histogram. Assuming L_c uniform containers in the range of $[-D_{\max}, D_{\max}]$ (usually $D_{\max} \approx 1$), the center of each container in the histogram is given by

$$z_n = D_{\max}\left(\frac{2n+1}{L_c} - 1\right), \quad n = 0, \ldots, L_c - 1. \tag{14}$$

The histogram $Q(n)$ is calculated by summing all the weighting factors of the time-frequency points with DOA values lying in each of the L_c containers previously defined.

$$Q(n) = \sum_{(k,m)\in\mathcal{S}_n} g(k,m), \quad n = 0, \ldots, L_c - 1, \tag{15}$$

where $g(k,m)$ is the weighting factor for the (k,m) point and $\mathcal{S}_n = \{(k,m) | L_c | D(k,m) - z_n | \langle 1, \Phi(k,m) \rangle \Phi_T\}$, i.e., the set of coherence-selected points with $D(k,m)$ in the value range defined by container n. The weighting factor is simply the mean amplitude value of that point in the two observation channels

$$g(k,m) = \frac{|X_1(k,m)| + |X_2(k,m)|}{2}. \tag{16}$$

Figure 5 shows different normalized histograms for the example mixture of three sources and different amounts of reverberation simulated with different wall reflection factors ρ.[29] It can be clearly observed how the peaks corresponding to the presence of the three speakers are clearly distinguishable in the anechoic case, but they broaden and overlap with one another as soon as reverberation increases (dotted lines). However, note how the peaky structure is highly enhanced when the histogram is computed from high-coherence time-frequency bins selected by means of a coherence test (solid line).

C. Multi-level thresholding

Thresholding is an important technique for image segmentation, which is used for the identification and extraction of targets from their background on the basis of the distribution of pixel intensities in image objects. In this work, we propose to consider the DOA map as an image where different objects are present (sources with different DOA). One of the most referenced thresholding methods in image processing is the Otsu method.[14] Several works have shown that, despite it was published a long time ago (1979), the Otsu method is one of the better threshold selection algorithms for general real world images.[30,31] The key concept underlying this segmentation method is the maximization of the *interclass variance* (or minimization of the *intraclass variance).* This clustering technique has been selected in the proposed audio separation framework for some remarkable advantages. On the one hand, it is very general; no specific histogram shape is assumed. This is a desirable feature because the shape of the histogram may change a lot depending on the mixing configuration, the spectral overlap, or the room reverberation. On the other hand, the Otsu method has been shown to work well and reliably under many different situations: different types of images, additive noise, etc. In addition, the extension of this algorithm to multi-level thresholding is straightforward, which enables to carry out the separation of more sources than mixtures. There also exists an efficient implementation of this algorithm, which opens the possibility to implement a real-time separation system. Moreover, as opposed to other clustering methods, there is no need for detecting peaks nor for specifying initial centroid values, which leads to a more unsupervised approach.

In our separation context, the aim of the proposed approach is to find the thresholds that maximize the interclass variance of the distribution of DOA estimates, given the histogram generated in the previous step. The ideas underlying the Otsu threshold selection method are described next.

1. The Otsu algorithm

In this section, the Otsu algorithm[14] is briefly described. The probability of a value in the middle of container n of the histogram is given by

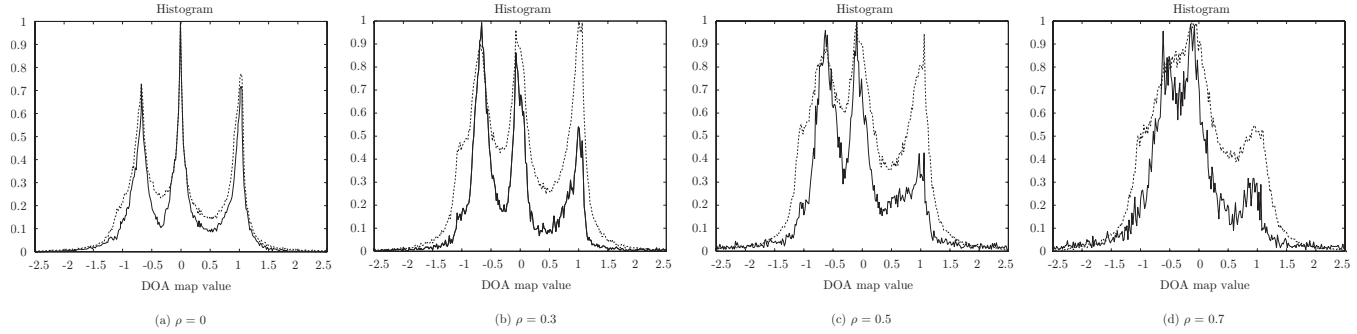


FIG. 5. Weighted histograms (normalized) for a mixture of three sources without pre-selection (dotted line) and with pre-selection (solid line) for different wall reflection factor ρ.

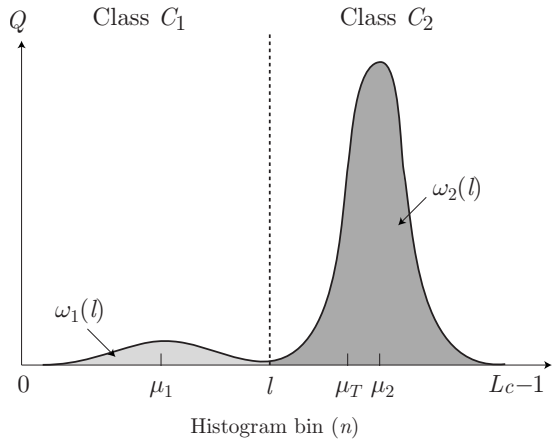


FIG. 6. Histogram divided into two classes.

$$p_n = \frac{Q(n)}{\kappa} \quad \text{with} \quad \kappa = \sum_{n=1}^{L_c} Q(n). \tag{17}$$

In the case of bi-level thresholding, the time-frequency points are divided into two classes: C_1, with DOA values in the range given by the histogram bins $n \in [1, \ldots, l]$, and C_2, with values within the bins $n \in [l+1, \ldots, L_c]$. The different variables involved in the segmentation are depicted in Fig. 6. The probability distributions for the two classes are as follows:

$$C_1: p_1/\omega_1(l), \ldots, p_l/\omega_1(l), \tag{18}$$

$$C_2: p_{l+1}/\omega_2(l), p_{l+2}/\omega_2(l), \ldots, p_{L_c}/\omega_2(l), \tag{19}$$

where $\omega_1(l) = \Sigma_{n=1}^{l} p_n$ and $\omega_2(l) = \Sigma_{n=l+1}^{L_c} p_n$.

The means for classes C_1 and C_2 are

$$\mu_1 = \sum_{n=1}^{l} n \frac{p_n}{\omega_1(l)}, \quad \mu_2 = \sum_{n=l+1}^{L} n \frac{p_n}{\omega_2(l)}. \tag{20}$$

Let μ_T be the mean value for all the selected DOA estimates. Then

$$\omega_1 \mu_1 + \omega_2 \mu_2 = \mu_T, \tag{21}$$

$$\omega_1 + \omega_2 = 1. \tag{22}$$

The between-class variance is defined as

$$\sigma_B^2 = \omega_1(\mu_1 - \mu_T)^2 + \omega_2(\mu_2 - \mu_T)^2. \tag{23}$$

For bi-level thresholding, it has been verified that the optimal threshold l^* is chosen so that the between-class variance σ_B^2 is maximized, that is

$$l^* = \arg \max_l \{\sigma_B^2(l)\}, \quad 1 \le l \le L_c. \tag{24}$$

The previous formula can be extended to multi-level thresholding as follows: assuming that there are $M-1$ thresholds, $\{l_1, l_2, \ldots, l_{M-1}\}$, which divide the DOA map into M classes: C_1 for $[1, \ldots, l_1]$, C_2 for $[l_1+1, \ldots, l_2]$, C_ϵ for $[l_{\epsilon-1}+1, \ldots, l_\epsilon]$, and C_M for $[l_{M-1}+1, \ldots, L_c]$. The optimal thresholds $l_1^*, l_2^*, \ldots, l_{M-1}^*$ are chosen by maximizing σ_B^2 as follows:

TABLE I. Computational complexity of Liao's fast algorithm.

$\sigma_B'^2$	Total computation
Addition	$(M-1)(L_c-M+1)^{M-1}2L_c$
Subtraction	$L_c(L_c-1)$
Multiplication	L_c
Division	$L_c(L_c+1)$
Direct index	$M(L_c-M+1)^{M-1}$

$$\{l_1^*, l_2^*, \ldots, l_{M-1}^*\} = \arg \max_{l_1, l_2, \ldots, l_{M-1}} \{\sigma_B^2(l_1, l_2, \ldots, l_{M-1})\}, \tag{25}$$

for $1 \le l_1 \le l_2 \cdots l_{M-1} < L_c$ and

$$\sigma_B^2(l_1, l_2, \ldots, l_{M-1}) = \sum_{\epsilon=1}^{M} \omega_\epsilon (\mu_\epsilon - \mu_T)^2, \tag{26}$$

with $\omega_\epsilon = \Sigma_{n \in C_\epsilon} p_n$, and $\mu_\epsilon = \Sigma_{n \in C_\epsilon} n(p_n/\omega_\epsilon)$.

The ω_ϵ in Eq. (26) is regarded as the zeroth-order cumulative moment of the ϵth class C_ϵ, and the numerator in the definition of μ_ϵ is regarded as the first-order cumulative moment of C_ϵ, that is

$$\mu(\epsilon) = \sum_{n \in C_\epsilon} n p_n. \tag{27}$$

Regardless of the number of classes being considered during the thresholding process, the sum of the cumulative probability functions of M classes equals one ($\Sigma_{\epsilon=1}^{M} \omega_\epsilon = 1$), and the mean of the DOA values considered is equal to the sum of the means of M classes weighted by their cumulative probabilities ($\mu_T = \Sigma_{\epsilon=1}^{M} \omega_\epsilon \mu_\epsilon$). Thus, the interclass variance in Eq. (26) can be rewritten as

$$\sigma_B^2(l_1, l_2, \ldots, l_{M-1}) = \sum_{\epsilon=1}^{M} \omega_\epsilon \mu_\epsilon^2 - \mu_T^2. \tag{28}$$

Because the second term in Eq. (28) is independent from the choice of the thresholds, the optimal thresholds can be chosen by maximizing $\sigma_B'^2$, which is defined as the summation term on the right-hand side of Eq. (28)

$$\{l_1^*, l_2^*, \ldots, l_{M-1}^*\} = \arg \max_{l_1, l_2, \ldots, l_{M-1}} \left\{ \sigma_B'^2 = \sum_{\epsilon=1}^{M} \omega_\epsilon \mu_\epsilon^2 \right\}, \tag{29}$$

The final thresholding values will be those in the middle of containers $n = l_\epsilon^*$:

$$\mathrm{Th}_\epsilon = z_n|_{n=l_\epsilon^*}. \tag{30}$$

2. Efficient implementation

Liao *et al.*[32] proposed a complexity-reduced algorithm based on the recursive calculation of $\sigma_B'^2$. By using a set of look-up tables, the complexity of the algorithm is reduced to the number of operations shown in Table I. It can be observed that the number of operations depends on the number of histogram bins L_c and the number of classes considered M, which is assumed to be equal to the number of sources N. If the number of sources is not known beforehand, several

methods can be used for estimating the number of sources or deciding the optimum number of thresholds.[33,34]

D. Binary masks

The calculated thresholds define the binary masks corresponding to each class in the DOA map. The separation binary masks are given by

$$M_j(k,m) = \begin{cases} 1 & \text{if } \mathrm{Th}_{j-1} < D(k,m) \le \mathrm{Th}_j \\ & \forall\, k,m \\ 0 & \text{elsewhere} \end{cases}, \tag{31}$$

with $j=1,\ldots,N$, N being the number of sources, $\mathrm{Th}_0=-1$, and $\mathrm{Th}_N=1$. The estimates of the images of the sources in each of the observation channels can be directly obtained by applying the separation masks to the STFT of the mixtures, as in Eq. (7).

1. Dynamic separation

Note that a dynamic approach of the described processing is straightforward. In Secs. III A–III C, a DOA map has been introduced and several thresholds have been calculated, considering a histogram that covered estimates over a set of time frames. However, the time length of the DOA map can be modified in real-time applications, allowing to recalculate the separating thresholds in a dynamic way. This approach opens the possibility of separating moving sources. In Sec. IV, several experiments will be carried out to evaluate the performance of the algorithm under different mixing situations and recording setups, including the case of moving speakers.

IV. EXPERIMENTS

In this section, we evaluate the performance of the proposed method under different mixing situations. The validity of the approach is shown by means of a set of experiments designed with the aim of discussing the performance of the algorithm in cases of special interest, involving different angular separation and number of speakers, different degrees of reverberation, moving sources, and speech sources recorded in a real room.

The input signals and processing parameters used in the experiments are as follows: the audio sources (male and female speeches) were extracted from the public data provided in *The 2008 Signal Separation Evaluation Campaign*.[35] They are sampled at 16 kHz, 16 bits, and have a duration of 10 s. The STFT was computed using Hann windows[36] of 1024 samples of length, with a hop size of 512 samples (50% overlap). The parameters of the Otsu algorithm were $L_c=200$, $\Phi_T=0.95$, $\lambda=0.6$, and $d=0.02$.

A. Objective performance criteria

Some performance measures in source separation processing have already been described in the literature. Vincent *et al.*[37] proposed some objective performance measures that are broadly used by the separation community. These measures are the source to distortion ratio (SDR), source to interferences ratio (SIR), and source to artifact ratio (SAR). In the *First Stereo Audio Source Separation Evaluation Campaign*,[38] the image to spatial distortion ratio (ISR) was introduced for taking into account the comparison with the true source images in multi-channel separation algorithms. These criteria can be computed for all types of separation algorithms, avoiding the need for knowing the separating filters or masks. The criteria derive from the decomposition of an estimated source image as

$$\hat{s}_{ij}(t) = s_{ij}(t) + e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t) + e_{ij}^{\mathrm{artif}}(t), \tag{32}$$

where $s_{ij}^{\mathrm{img}}(t)$ is the true source image, and $e_{ij}^{\mathrm{spat}}(t)$, $e_{ij}^{\mathrm{interf}}(t)$, and $e_{ij}^{\mathrm{artif}}(t)$ are distinct error components representing spatial (or filtering) distortion, interference, and artifacts. The signal $s_{ij}(t)+e_{ij}^{\mathrm{spat}}(t)$ is obtained by least-squares projection of the estimated source image onto the signal subspace spanned by filtered versions of the true source image. Similarly, the signal $s_{ij}(t)+e_{ij}^{\mathrm{spat}}(t)+e_{ij}^{\mathrm{interf}}(t)$ is obtained by least-squares projection of the estimated source image onto the signal subspace spanned by filtered versions of all the source images.

From this decomposition, the mentioned performance measures are defined as follows: the image to spatial distortion ratio is given by

$$\mathrm{ISR} = 10\log_{10}\frac{\sum_{i=1}^{2}\sum_{t} s_{ij}(t)^2}{\sum_{i=1}^{2}\sum_{t} e_{ij}^{\mathrm{spat}}(t)^2}. \tag{33}$$

Similarly, we define the source to interferences ratio as

$$\mathrm{SIR} = 10\log_{10}\frac{\sum_{i=1}^{2}\sum_{t} (s_{ij}(t) + e_{ij}^{\mathrm{spat}}(t))^2}{\sum_{i=1}^{2}\sum_{t} e_{ij}^{\mathrm{interf}}(t)^2}, \tag{34}$$

the source to artifact ratio as

$$\mathrm{SAR} = 10\log_{10}\frac{\sum_{i=1}^{2}\sum_{t} (s_{ij}(t) + e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t))^2}{\sum_{i=1}^{2}\sum_{t} e_{ij}^{\mathrm{artif}}(t)^2}, \tag{35}$$

and finally, the signal to distortion ratio as

$$\mathrm{SDR} = 10\log_{10}\frac{\sum_{i=1}^{2}\sum_{t} s_{ij}(t)^2}{\sum_{i=1}^{2}\sum_{t} (e_{ij}^{\mathrm{artif}}(t) + e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t))^2}. \tag{36}$$

Recall again that the above definitions were proposed by Vincent[38] with the aim of extending his original work[37] to multi-channel signals.

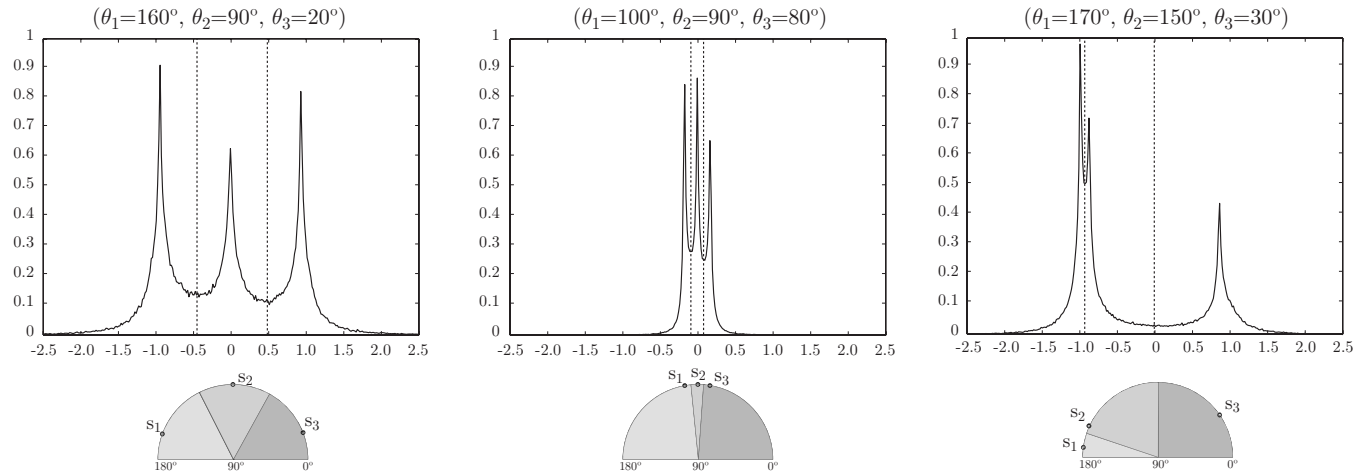


FIG. 7. Histograms and separation thresholds for different source arrangements with ρ=0 (anechoic). The angular segmentation of the azimuth plane for each case is depicted at the bottom of each graph, showing how the sources are effectively clustered in different angular sections even when they are close to each other.

B. Angular separation of speakers

The capability of the algorithm to decide the angular areas wherein the speakers are located is shown in Fig. 7, where an anechoic scenario and three speech sources have been considered in order to better visualize how the method works (later experiments will consider room reflections and more speakers). The separation thresholds (marked as dotted vertical lines) are accurately positioned between the source peaks, segmenting the azimuth plane into angular areas that denote the presence of each speaker. Notice how the segmentation is successfully performed even in the case when the angular separation between the sources is very small (10°).

C. Robustness against reverberation

The degradation introduced by reverberation with respect to the anechoic case is shown in Fig. 8. The peaks that appear in the histogram are more distorted when the wall reflection factor ρ is increased, making the detection of the speakers more difficult in the presence of room reflections. Note that despite the observed degradation, only slight differences between the angular source areas are produced. It is important to remark that the histograms have been constructed considering only high-coherence bins, so the effect that pre-selection has in improving the robustness against reverberation is here clearly observed.

D. Number of speakers

The number of speakers in the mixture affects the performance of the algorithm. Nevertheless, Fig. 9 shows how the thresholds correctly separate the source distributions even when four sources are present and room reflections are considered. When N=5 two sources are clustered into the same region, but the other sources have been correctly detected. Therefore, even in the case when many sources are present, the algorithm is still able to provide correct results for most of the sources.

E. Moving sources

The dynamic approach of Sec. III D 1 has been considered in an example, where there are two static sources and

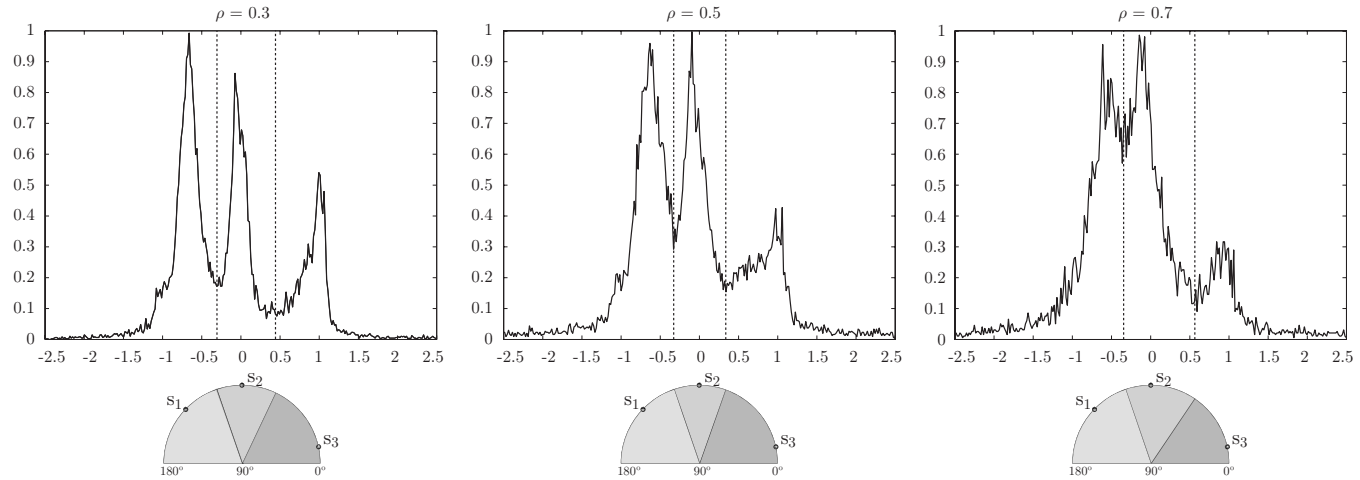


FIG. 8. Histograms and separation thresholds for different degrees of reverberation as a function of the wall reflection factor ρ. The speakers are at locations ((θ_1=135°, θ_2=90°, and θ_3=20°)). Note that despite the degradation of the histograms with reverberation, only slight differences between the angular source areas are produced.

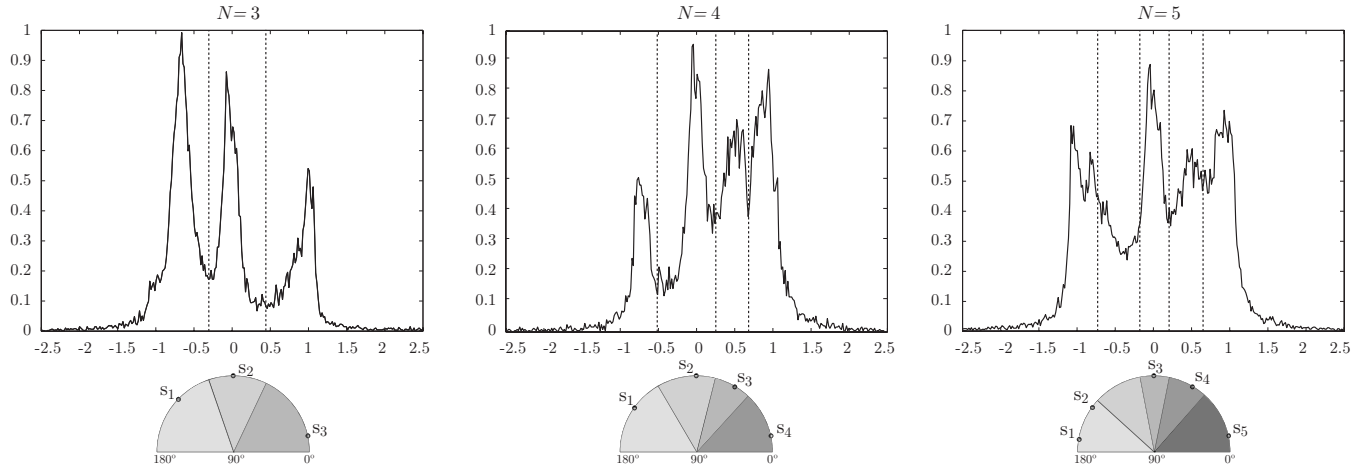


FIG. 9. Histograms and separation thresholds for different number of speakers in a room with moderate reverberation $\rho=0.3$: (a) $\theta_1=135°$, $\theta_2=90°$, and $\theta_3=20°$; (b) $\theta_1=135°$, $\theta_2=90°$, $\theta_3=60°$, and $\theta_4=20°$; (c) $\theta_1=170°$, $\theta_2=135°$, $\theta_3=90°$, $\theta_4=60°$, and $\theta_5=20°$.

one moving source going from $\theta=0°$ to $\theta=150°$ in a time period of 4 s. The thresholds are calculated every 0.8 s and the separation masks are consequently updated, taking into account the new histograms. Figure 10 shows the histograms obtained in each update step with the thresholds obtained as output of the segmentation process. Note how the thresholds are correctly updated in each step. At $t=2.4$ s the moving source is at the same location as one of the static sources. Although the angular distributions are correctly identified, the overlapping sources cannot be separated if they are at the same angular location. Obviously, when there are crossings between angular source positions, the separation masks are interchanged; thus, a tracking stage must be included in order to preserve the identity between each source and its corresponding mask. The solutions to these problems are out of the scope of this paper, but will be addressed in the future.

F. Source separation performance

The objective performance measures for different separation examples with $N=3$ and $N=4$ sources are shown in Tables II and III, respectively. The values given correspond to the average results over all the sources for each separation experiment. The angular positions of the sources were ($\theta_1=135°$, $\theta_2=90°$, and $\theta_3=20°$) in the three source example, and ($\theta_1=135°$, $\theta_2=90°$, $\theta_3=60°$, and $\theta_3=20°$) in the four source case. For comparison purposes, the performance measures achieved by the ideal binary mask in each experiment are also included. Consider that the results provided by the ideal binary mask are always the optimal and can be understood as the maximum quality achievable by binary time-frequency masking. However, the original sources are needed to compute for the ideal binary mask, and state-of-

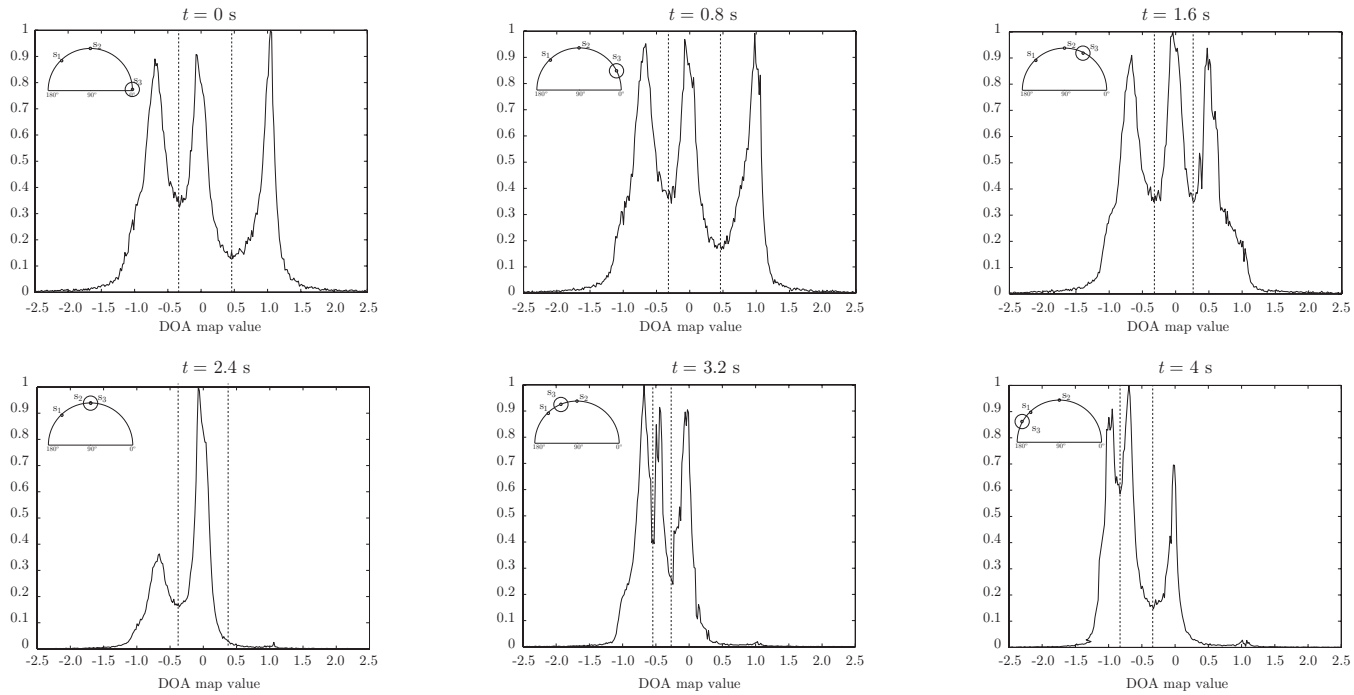


FIG. 10. Histograms and separation thresholds for the moving source example at different time instants.

TABLE II. Average objective performance measures for $N=3$ sources.

	$N=3$							
	$\rho=0$		$\rho=0.3$		$\rho=0.5$		$\rho=0.7$	
	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal
SDR	6.4	9.5	2.6	9.2	1.9	8.9	0.9	8.2
ISR	15.1	20.0	10.1	18.9	7.8	18.2	5.6	17.4
SIR	15.3	22.0	10.7	20.6	8.3	19.7	2.8	18.3
SAR	6.9	9.7	5.6	9.6	4.6	9.3	4.4	8.5

the-art algorithms are still far from achieving similar performance for reverberant scenarios and a high number of sources. As expected, the performance of the algorithm is decreased when there are more sources in the mixture or the degree of reverberation is higher. This is not only true for the proposed algorithm but also for the case of the ideal binary mask, which also shows a degradation in the demixing performance due to these two factors. Of all the performance measures, SAR and SDR are more affected by an increment in the number of sources, while reverberation has a greater influence on the SIR and ISR measures.

G. Separation performance in real office rooms

Experiments using mixture signals recorded in a real office room have also been conducted in the context of the 2008 Signal Separation Evaluation Campaign. These were acquired in a chamber with cushion walls, using the loudspeaker and microphone arrangement depicted in Fig. 11. The room had a reverberation time of $RT_{60}=130$ ms, and the experiments were carried out by considering mixtures of male and female speech sources. These mixtures were sampled at 16 kHz and had a duration of 10 s. The intermicrophone distance was set to 5 cm. Table IV shows the average results for these experiments together with the results of the ideal binary mask. Note how the results in a real case are even better than those obtained in previous simulations with reverberation. Audio files can be downloaded at ⟨www.irisa.fr/metiss/SiSEC08/SiSEC_underdetermined/test_eval.html⟩. The measures shown in the table confirm how the method is able to separate real mixtures of speech with little computation time. This is further discussed in Sec. IV H.

H. Computation time

The times required by the proposed algorithm for deciding the angular separation areas as a function of the number of sources N and the number of histogram bins L_c are shown in Table V. The processor used for measuring these times was an Intel Core2 2.0 GHz. Note how the number of sources considered has the greatest influence on the computation time. Nevertheless, a very high number of histogram bins is not necessary, since the angular resolution of the method is mainly imposed by reverberation. Moreover, L_c can be modified, depending on the computational power of the processor, which is another advantage for real-time implementations.

TABLE III. Average objective performance measures for $N=4$ sources.

	$N=4$							
	$\rho=0$		$\rho=0.3$		$\rho=0.5$		$\rho=0.7$	
	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal
SDR	3.7	8.1	1.5	7.9	0.1	7.3	−0.1	6.9
ISR	11.1	17.8	7.3	17.2	5.3	16.5	3.4	15.4
SIR	10.6	20.3	5.8	19.5	3.5	18.3	1.7	17.1
SAR	3.7	8.1	3.4	8.1	2.3	7.8	2.1	6.5

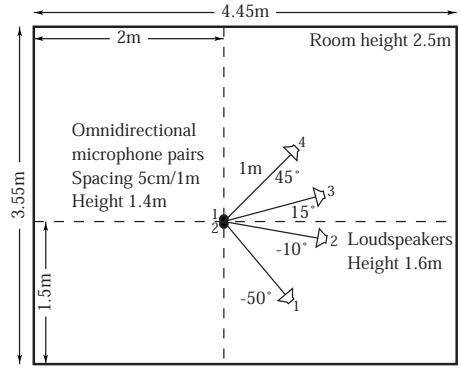


FIG. 11. Arrangement for the recording of real speech mixtures (extracted from Ref. 38).

V. CONCLUSION

In this paper, a method for separating real mixtures of speech using two microphones has been presented. The proposed algorithm is based on grouping time-frequency points with similar DOA using a multi-level thresholding approach. The thresholds are calculated via the maximization of the interclass variance between DOA estimates, and they allow to identify angular sections, wherein the speakers are located with a strong likelihood.

Separation is efficiently achieved without the need for estimating the mixing matrix nor detecting peaks or specifying initial centroid values as with other clustering approaches. Moreover, although the method is based on the anechoic mixing model, two features have been proposed to improve the robustness against reverberation: a coherence test and a weighted histogram. These features allow to perform separation in realistic acoustic scenarios.

The validity of the method has been studied over a complete set of experiments, including both simulations and real speech mixtures of up to four sources. The results show that

TABLE IV. Average objective performance measures for recordings in a real office room.

	$N=3$				$N=4$			
	Female		Male		Female		Male	
	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal	Proposed	Ideal
SDR	2.4	12.1	2.2	9.6	2.2	8.4	2.1	7.6
ISR	6.5	21.8	6.8	18.3	6.6	16.0	6.1	15.5
SIR	6.4	22.6	5.9	20.2	6.5	17.4	6.6	17.5
SAR	7.8	12.8	5.2	9.9	5.5	9.0	4.5	7.8

TABLE V. Computation times for different number of histogram bins L_c and number of sources N.

	Computing times (ms)					
L_c	10	20	100	200	300	500
$N=2$	1.2	1.2	1.8	3.3	7.7	19.0
$N=3$	1.5	1.5	2.0	4.4	10.1	24.4
$N=4$	1.6	1.6	9.2	52.2	164.2	743.4

the proposed algorithm provides state-of-the-art performance with a low computation time, which makes it very appropriate for real-time speech applications.

ACKNOWLEDGMENTS

The Spanish Ministry of Science and Innovation supported this work under Grant Nos. TEC2009-14414-C03-01 and TEC2006-13883-C04-01.

[1]E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979 (1953).
[2]A. W. Bronkhorst, "The cocktail party phenomenom: A review of research on speech intellibility in multiple-talker conditions," Acustica **86**, 117–128 (2000).
[3]N. Roman, S. Srinivasan, and D. Wang, "Binaural segregation in multisource reverberant environments," J. Acoust. Soc. Am. **120**, 4040–4051 (2006).
[4]J. F. Cardoso, "Blind signal separation: Statistical principles," Proc. IEEE **86**, 2009–2025 (1998).
[5]S. Pedersen, J. Larsen, U. Kjems, and L. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook of Speech Processing* (Springer, New York, 2007).
[6]H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," IEEE Trans. Speech Audio Process. **12**, 530–538 (2004).
[7]P. Bofill and M. Zibulevski, "Underdetermined blind source separation using sparse representations," Signal Process. **81**, 2353–2362 (2001).
[8]S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput. (USA) **20**, 33–61 (1998).
[9]M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," Neural Comput. **12**, 337–365 (2000).
[10]D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," Proc. Natl. Acad. Sci. U.S.A. **100**, 2197–2202 (2003).
[11]S. Rickard and O. Yilmaz, "On the W-disjoint orthogonality of speech," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL (2002), pp. 529–532.
[12]O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," IEEE Trans. Signal Process. **52**, 1830–1847 (2004).
[13]S. Araki, H. Sawada, R. Mukai, and S. Makino, "A novel blind source separation method with observation vector clustering," in Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '05), Eindhoven (2005), pp. 117–120.
[14]N. Otsu, "A threshold selection method from gray-level histogram," IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979).
[15]M. Cobos and J. J. Lopez, "Stereo audio source separation based on time-frequency masking and multilevel thresholding," Digit. Signal Process. **18**, 960–976 (2008).
[16]*Blind Speech Separation*, edited by S. Makino, T. W. Lee, and H. Sawada (Springer, New York, 2007).
[17]L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[18]J. J. Burred and T. Sikora, "On the use of auditory representations for sparsity-based sound source separation," in Proceedings of the Fifth International Conference on Information, Communications and Signal Processing (ICICS 2005), Bangkok, Thailand (2005).
[19]A. Jourjine, S. Richard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '00, Turkey (2000), Vol. **5**, pp. 2985–2988.
[20]M. Cobos and J. J. Lopez, "Improving isolation of blindly separated sources using time-frequency masking," IEEE Signal Process. Lett. **15**, 617–620 (2008).
[21]*Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, edited by D. L. Wang and G. J. Brown (Wiley, New York, 2006).
[22]D. Wang, "Time frequency masking for speech separation and its potential for hearing aid design," Trends Amplif. **12**, 332–353 (2008).
[23]C. Faller, "Parametric coding of spatial audio," Ph.D. thesis, École Polytechynique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (2004).
[24]C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089 (2004).
[25]S. Mohan, M. E. Lockwood, M. L. Kramer, and D. L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test," J. Acoust. Soc. Am. **123**, 2136–2147 (2008).
[26]C. Avendano and J. M. Jot, "Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02), Orlando, FL (2002).
[27]G. C. Carter, "Coherence and time delay estimation," Proc. IEEE **75**, 236–255 (1987).
[28]H. F. Silverman and J. M. Sachar, "The time-delay graph and the delayogram—New visualizations for time delay," IEEE Signal Process. Lett. **12**, 301–304 (2005).
[29]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[30]P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," Comput. Vis. Graph. Image Process. **41**, 233–260 (1988).
[31]M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," J. Electron. Imaging **13**, 146–165 (2004).
[32]P. Liao, T. Chen, and P. Chung, "A fast algorithm for multilevel thresholding," J. Inf. Sci. Eng. **17**, 713–717 (2001).
[33]P. Y. Yin and L. H. Chen, "A fast iterative scheme for multilevel thresholding methods," Signal Process. **60**, 305–313 (1997).
[34]B. Loesch and B. Yang, "Source number estimation and clustering for underdetermined blind source separation," in Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC), Seattle, WA (2008).
[35]E. Vincent, S. Araki, and P. Bofill, "The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation," Lect. Notes Comput. Sci. **5441**, 734–741 (2009).
[36]R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (Dover, New York, 1958).
[37]E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," IEEE Trans. Speech Audio Process. **14**, 1462–1469 (2006).
[38]E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. P. Rosca, "First stereo audio source separation evaluation campaign: Data, algorithms and results," in Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA 2007), London, UK (2007).

Do ferrets perceive relative pitch?

Pingbo Yin, Jonathan B. Fritz, and Shihab A. Shamma
Neural Systems Laboratory, Institute for Systems Research, University of Maryland, College Park, Maryland 20742

(Received 18 March 2009; revised 13 November 2009; accepted 13 December 2009)

The existence of relative pitch perception in animals is difficult to demonstrate, since unlike humans, animals often attend to absolute rather than relative properties of sound elements. However, the results of the present study show that ferrets can be trained using relative pitch to discriminate two-tone sequences (rising vs. falling). Three ferrets were trained using a positive-reinforcement paradigm in which sequences of reference (one to five repeats) and target stimuli were presented, and animals were rewarded only when responding correctly to the target. The training procedure consisted of three training phases that successively shaped the ferrets to attend to relative pitch. In Phase-1 training, animals learned the basic task with sequences of invariant tone-pairs and could use absolute pitch information. During Phase-2 training, in order to emphasize relative cues, absolute pitch was varied each trial within a two-octave frequency range. In Phase-3 training, absolute pitch cues were removed, and only relative cue information was available to solve the task. Two ferrets successfully completed training on all three phases and achieved significant discriminative performance over the trained four-octave frequency range. These results suggest that ferrets can be trained to discern the relative pitch relationship of a sequence of tone-pairs independent of frequency. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290988]

PACS number(s): 43.80.Lb, 43.66.Gf [MJO] Pages: 1673–1680

I. INTRODUCTION

A significant amount of information is encoded in the contour patterns (rises and falls) of the pitch of acoustic signals, such as in speech and music. For example, human subjects can easily recognize sentence type on the basis of pitch contour alone in the absence of other information (Ladefoged, 1982). The frequency transpositions of a melody are readily recognized by most adult, and even infant, human listeners as the "same" and are perceived as structural equivalents of the original melody (Dowling and Fujitani, 1971; Demany and Armand, 1984; Trehub *et al.*, 1984; Trehub and Hannon, 2006). Although human listeners can remember the exact musical intervals of familiar melodies, they appear to remember only the frequency contour of less familiar or novel stimuli (Dowling and Fujitani, 1971; Dowling, 1978; Bartlett and Dowling, 1980). Unlike humans, who attend chiefly to the relationships *between* sound elements, animals more heavily weight the *absolute* frequency of sound elements in their perceptual decisions and appear to be less attentive to *relative* pitch changes. Consequently, it has been difficult to train animals to attend to the relative pitch between sound elements, as D'Amato (1988) concluded after extensive behavioral research on monkeys and rats.

Most studies conducted with nonhuman species, including several species of birds (Hulse and Cynx, 1985, 1986; Ratcliffe and Weisman, 1986; Dooling *et al.*, 1987; Page *et al.*, 1989; Cynx, 1993; Weisman *et al.*, 2004) and monkeys (D'Amato and Salmon, 1984; D'Amato, 1988; Izumi, 2001, 2003; Brosch *et al.*, 2004, 2006), suggest that animals generally encode absolute pitch and have rather limited abilities to recognize the relative pitch contours of tonal stimuli. Songbirds have been shown to learn a relative pitch strategy, recognizing an ordinal rule for tone sequences that rise or fall in frequency regardless of the absolute frequency components. However, when the sequences were shifted out of the trained frequency range, they lost the discrimination. It then required as many trials to acquire a new discrimination as they needed to learn in the original discrimination (Hulse and Cynx, 1985; Cynx, 1993). Furthermore, songbirds failed to learn relative pitch discrimination when the absolute pitch cues were removed from the training (Page *et al.*, 1989), indicating the primacy of absolute pitch perception in these species. Although a frequency range constraint was also noted in nonhuman mammals (Izumi, 2001, 2003), there are now two studies indicating that some nonhuman mammalian species are capable of relative pitch as measured by octave generalizations—rhesus monkeys (Wright *et al.*, 2000) and dolphins (Ralston and Herman, 1995).

In a recent behavioral study, Walker *et al.* (2009) successfully trained ferrets on a two-alternative forced choice task to discriminate sounds that were higher or lower in pitch than a reference sound. Since the reference sound remained constant throughout a given session, the animal could use absolute strategies to solve the task. However, the result might also suggest that ferrets can be trained to utilize relative pitch cues in sequential sounds. The goal of the present study was to develop a new animal model to study the neural mechanisms underlying auditory pattern categorization based on direction of pitch changes (pitch contours) of tone sequences, and more generally to understand the neural basis and correlates of recognition and discrimination of spectrotemporally complex sounds. A training procedure which gradually directed animals to attend to the relative pitch change of two-tone sequences (rising vs falling) has been successfully developed, and the present report provides evi-

dence of ferrets' capability to categorize tonal patterns solely on the basis of these two-tone step changes over the trained frequency range.

II. METHODS

A. Subjects

Three naïve female adult ferrets, weighing 600–900 g, were used in this behavioral study. The animals were trained on a positive-reinforcement operant paradigm with water as reward. The ferrets were placed on a water-control protocol on which they were typically trained 5 days per week and obtained *ad libitum* water over the weekend. On training days, animals received one or two training sessions (~100 trials in each session to satiation). All procedures conformed to the NIH policy on experimental animal care and use and were approved by the IACUC of the University of Maryland.

B. Experimental apparatus

Ferrets were tested in a customized-design wire mesh training cage (8 in. width × 15 in. depth × 9 in. height) which was placed within a single-walled, sound attenuated chamber (IAC). A lick sensitive waterspout (1 in. × 1.5 in.) stood 5 in. above the floor of the training cage. The waterspout was connected to a computer controlled water dispenser (Crist Instrument Co., Inc., Maryland, USA). A loudspeaker (Manger MSW, Germany) was positioned 10 cm in front of the cage, and the animal's behavior was monitored by video camera.

C. Basic behavioral paradigm

The positive-reinforcement operant paradigm used in these experiments differed substantially from previous behavioral studies in the laboratory that used a conditioning avoidance paradigm (Fritz *et al.*, 2003). In the current study, ferrets were trained to lick a waterspout as the behavioral response to a target sound (unlike the conditioned avoidance paradigm, in which ferrets learned to refrain from ongoing licking when the target sound was presented). Each training session started with delivery of a drop of water (~0.5 ml) to initiate licking of the waterspout. The first trial, and subsequent trials, began after the animal had consumed the water and then refrained from licking the spout for a minimum of 0.5 s, as illustrated in Fig. 1. A trial consisted of a sequence of one to five similar reference (non-target) sounds, followed by a different (target) sound. The inter-stimulus-interval between all references and target sound was 1.25 s. The animal was rewarded with a small drop of water when it licked the spout within a given time window *after* the target sound (*last* shaded area in Fig. 1). It received a 3–6 s timeout penalty if it did not lick during the target sound. The target reward drop volume (0.1–0.3 ml) was adjusted for each trial according to the licking pattern during the preceding reference stimuli. Specifically, the reward drop volume was inversely proportional to false-alarm rate of the trial in order to discourage licking during the reference stimuli. An additional click sound was played as a secondary "reinforcer" following water delivery during the early stages of training. The total

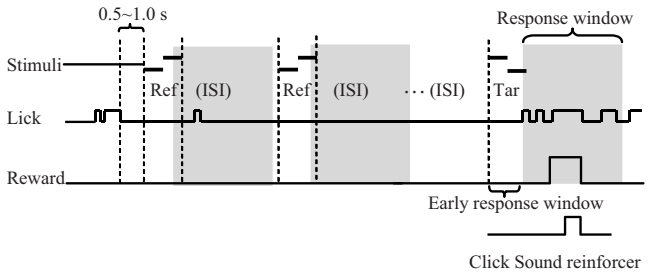


FIG. 1. *Positive-reinforcement operant paradigm.* A trial was initiated when the animal refrained from licking the waterspout for 0.5–1.0 s. A reference sound (non-target) was presented and repeated randomly one to five times after trial initiation. A target sound followed the reference sounds. When the ferret licked the waterspout within a 1.2 s response time window after target onset (*the shaded target period*), its response was counted as a *hit*, which was followed by water reward. If the ferret licked the waterspout within a corresponding time window after reference sound onset (*the shaded reference period*), its response was counted as a *false alarm*, which caused reduction in reward volume. A miss of the target lead to a 3–6 s timeout penalty after completion of the trial. A click sound was played as a secondary reward reinforcer after water reward delivery.

number of reference stimuli presented in given trial varied from trial to trial and was selected from a pseudo-random sequence, in which there was an equiprobable likelihood that the target sound would be presented at each position in the sequence (from second to fifth position). A training session ended when the animal did not lick the spout in two consecutive trials.

D. Training procedure and stimuli

Training began with a 1–2 day habituation period during which animals were allowed to explore the training cage and learned to obtain water by licking the waterspout. Training was continued by a pre-training phase of approximately 2 weeks in which water delivery was associated with sounds.

All acoustic stimuli were two-tone sequences, 300 ms in duration, and ~70 dB SPL. Each tone component in the two-tone sequence was 150 ms duration and was ramped with 5 ms rise-fall time. There was no silent gap between the tones. The frequency separation between the two tones was 1/3 (Phase 1) or 1/2 octaves (Phases 2 and 3).

Two ferrets (H and J) were trained to discriminate the downward sequences (*the target sequence*) from the upward sequences (*the reference sequence*), and one ferret (M) was trained to discriminate upward sequences (*the target sequence*) from the downward sequences (*the reference sequence*). Animals underwent a three-phase training schedule [Figs. 2(A)–2(C)] to be gradually directed to the final task requirements.

Initially, ferrets were trained with an easy version of the task (*Phase-1 training*), in which both the reference and target sequences were comprised of a fixed frequency tone-pair [the left panel in Fig. 2(D)]. The same tone-pair (with the same two tones arranged either upward or downward) was used during this entire training phase [Fig. 2(A)]. To perform the task, animals could either utilize the absolute frequencies of the tones (e.g., the initial and/or terminal pitch of the sequence) and/or relative properties of the tone sequence, specifically the direction of pitch change (rising vs falling).

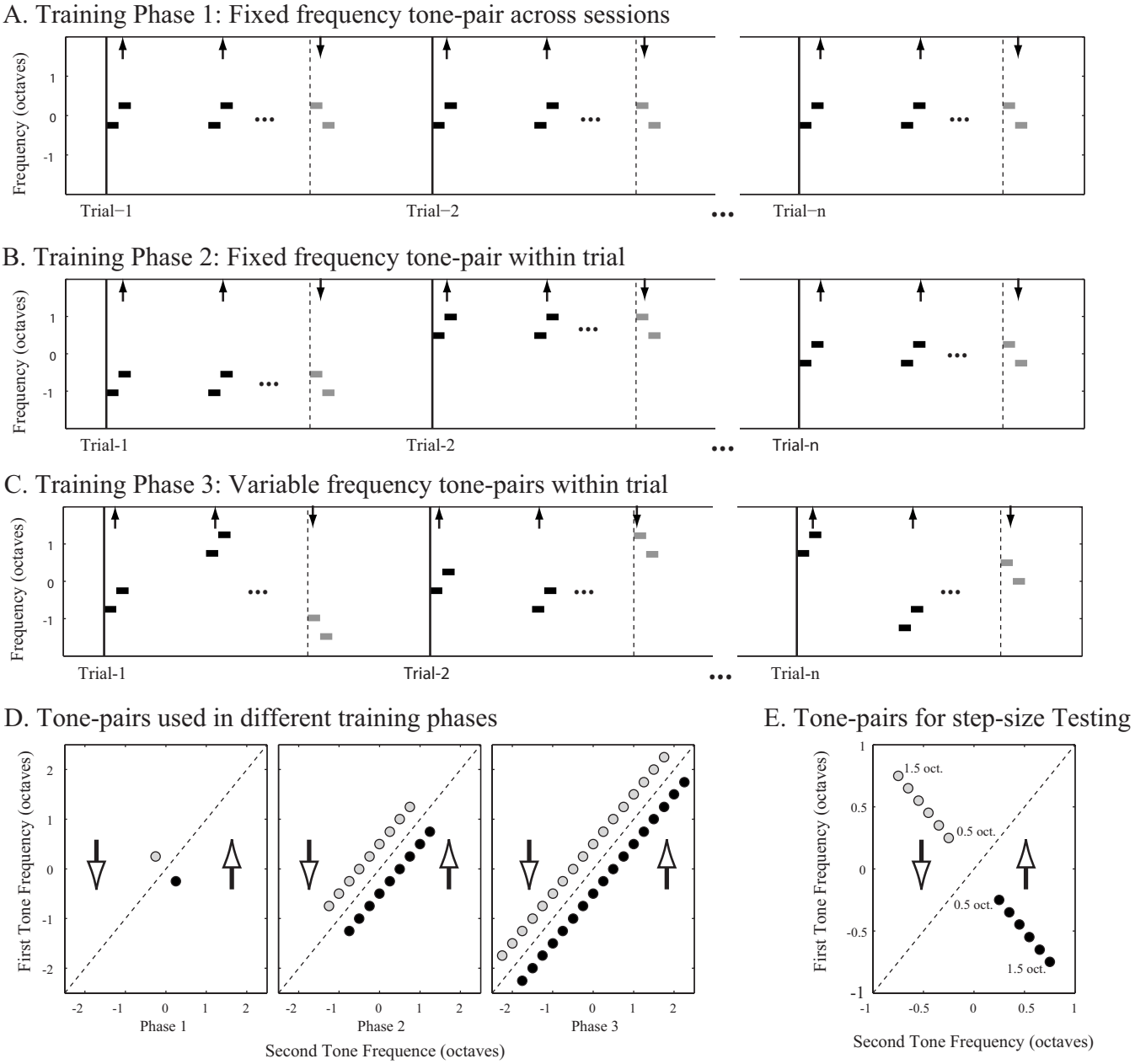


FIG. 2. *Two-tone sequence discrimination task: Training procedures and stimuli sets*. (A) Phase-1 training: The reference and target sequences were comprised of one tone-pair with fixed frequency across all training sessions. (B) Phase-2 training: The reference and target sequences for each trial were comprised of one tone-pair with fixed frequency, while the frequency of the tone-pair was changed randomly between successive trials. (C) Phase-3 training: Each of the reference or target sequences in a given trial was comprised of different tone-pairs for which the frequency of the tone-pair was chosen randomly from a set of 17 possible pairs. (D) Tone-pairs in different training phases: The rising and falling two-tone sequences were made from 1 tone-pair in Phase 1 (*the left panel*), 9 tone-pairs in Phase 2 (*the middle panel*), and up to 17 tone-pairs in Phase 3 (*the right panel*). The tone components of the two-tone sequence were $\frac{1}{2}$ octave apart in frequency, and the frequency contours of the sequences were roved up or down with $\frac{1}{4}$ octave increments. (E) Tone-pairs for step-size testing: The stimulus set varied the frequency separation between component tones, ranging from 0.5 up to 1.5 octaves. The solid vertical lines in (A)–(C) indicate the beginning of the trials. The vertical dashed lines indicate the target onset of each trial. The diagonal lines in (D) and (E) denote the iso-frequency line. The up and down arrows indicate directions of the reference and target sequence in (A)–(C) and of the quadrants in (D) and (E).

Once the behavioral criterion was achieved (see Sec. II E), two animals received one or more additional tone-pairs in Phase-1 training. All animals then began Phase-2 training, in which the two-tone sequence for each trial consisted of a randomly picked tone-pair within a limited frequency range [Fig. 2(B)]. The frequency of the tone-pair was varied between trials over a frequency range that was gradually expanded up to two octaves around the initial frequency with $\frac{1}{4}$ octave increments, forming nine different rising or falling sequences at the end of this phase [the middle panel in Fig. 2(D)]. In the final stage, training Phase 3, all reference and target sequences within each trial were randomly chosen [Fig. 2(C)] so that the absolute frequency of the sequences varied, and the only fixed parameter was the direction of pitch change (rising or falling). The frequency range of the tone-pairs at this training stage was expanded to four octaves for a total of 17 rising or falling sequences [the right panel in Fig. 2(D)].

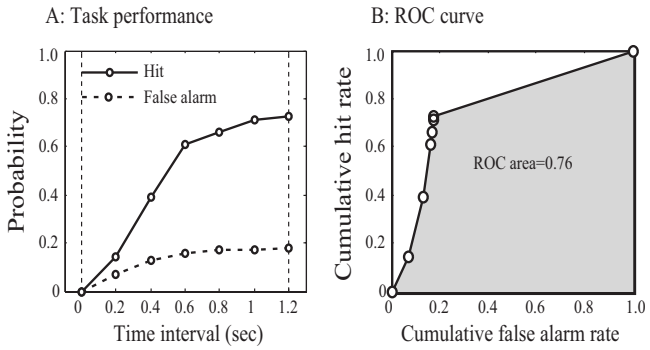


FIG. 3. *Construction of ROC curve.* (A) The probabilities for hit (*solid line*) and false alarm (*dashed line*) were independently computed at each of the time intervals from 0.0 to 1.2 s with 0.2 s increments following the onset of the response window after the target and reference. The vertical dashed lines indicate the start and the end of the response window. (B) The false-alarm probability function was plotted against hit probability function to construct the ROC curve (*solid line*). The area under the ROC curve (*shaded area*) was a measure of discriminative performance of the task.

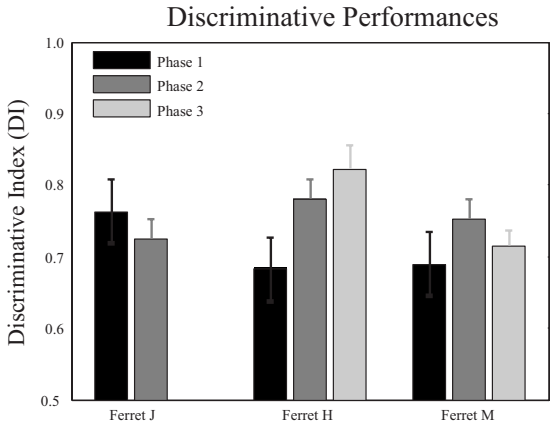


FIG. 4. *Discriminative performances for all training phases.* The bar plots show the average DI values, each of which was computed from ten consecutive sessions of Phase 1 (*black*), Phase 2 (*dark gray*), and Phase 3 (*light gray*) performances after reaching training criterion. The error bar indicates standard deviation.

A different stimulus set was used to test the effect of the step-size of tone-pairs on the discriminative performance after completion of Phase-3 training. This stimulus set included six upward and six downward sequences, which were compromised of six tone-pairs whose frequencies were [1200 1697], [1120 1819], [1045 1949], [975 2089], [909 2239], and [849 2400] Hz, respectively. Therefore, the frequency separation (or interval) between component tones in those sequences varied from 0.5 to 1.5 octaves [see Fig. 2(E)].

E. Data analysis

The timing of the first-lick after each of the references and target stimuli was recorded as the behavioral response to the stimulus on each trial. The task performance level was assessed by an analysis based on signal detection theory, in which both behavioral response accuracy and the behavioral response latency were exploited. The use of latency information in the analysis was useful in case of a difficult discrimination (Carterette *et al.*, 1965; Emmerich *et al.*, 1972) and also in obtaining a sufficient number of probability values for accurate determination of a receiver operating characteristic (ROC). A "first-lick" was defined as a hit or a false alarm depending on whether it was fallen in the response window following a target or non-target (the reference) sound (the shading zones after each sequence in Fig. 1). The probabilities of the hit (response after a target sound) and the false alarm (response after a non-target sound) were independently computed as function of the time intervals (from 0.0 to 1.2 s with an increment of 0.2 s) following the onset of the response window [Fig. 3(A)]. The ROC was then formed by the obtained probability functions (i.e., hit rate vs false-alarm rate function) [Fig. 3(B)]. The area under the ROC was taken as a measure of the task performance and was defined as the discriminative index (DI). This index yields a value of 0.5 for random performance, greater than 0.5 but less than 1.0 for nonrandom performance, and 1.0 for perfect performance. In each training session, the DI value was calculated from the original data and also from bootstrapped trials. In bootstrapped trials, the relationship between the behavioral responses (the First-lick time) and the stimulus tags (reference and target) was shuffled, and a shuffled-DI value was calculated. This process was repeated 50 times, and the mean value and the standard deviation of these shuffled-DI values were determined. A training session was considered to show significant discriminative performance if the obtained DI value was more than two standard deviations above the shuffled-DI mean. The behavioral criterion for achieving successful performance was defined as significant discriminative performance for a minimum of five consecutive training sessions. Animals could receive additional training on a given training phase after reaching criterion. Sessions with less than 40 trials were excluded from further analysis.

III. RESULTS AND DISCUSSIONS

Three ferrets were trained on the two-tone discrimination task. Two of them completed all three phases of training. The third ferret was suspended from further training after completion of two phases (it was withdrawn from the study because it was suffering weight loss due to a severe and debilitating gastrointestinal infection). Figure 4 shows the mean performance of the ferrets across ten consecutive sessions after reaching behavioral criterion at each of the three training phases. All three animals reached behavioral criterion in the first two training phases and yielded mean DI values between 0.68–0.76 ($M=0.71$, s.d.$=0.05$) for Phase 1 and 0.72–0.77 ($M=0.75$, s.d.$=0.03$) for Phase 2. The performance of the two animals completing the final training phase (Phase 3) yielded DI values of 0.71 and 0.82, respectively ($M=0.76$, s.d.$=0.08$).

A. Phase-1 training and frequency transposition

Training in Phase 1 took about 30–60 sessions before animals reached behavioral criterion. Since animals in Phase 1 could use either absolute or relative pitch to discriminate the target from reference sequences, if relative was being employed, then animals would easily generalize their training on the first tone-pair to a new tone-pair with different

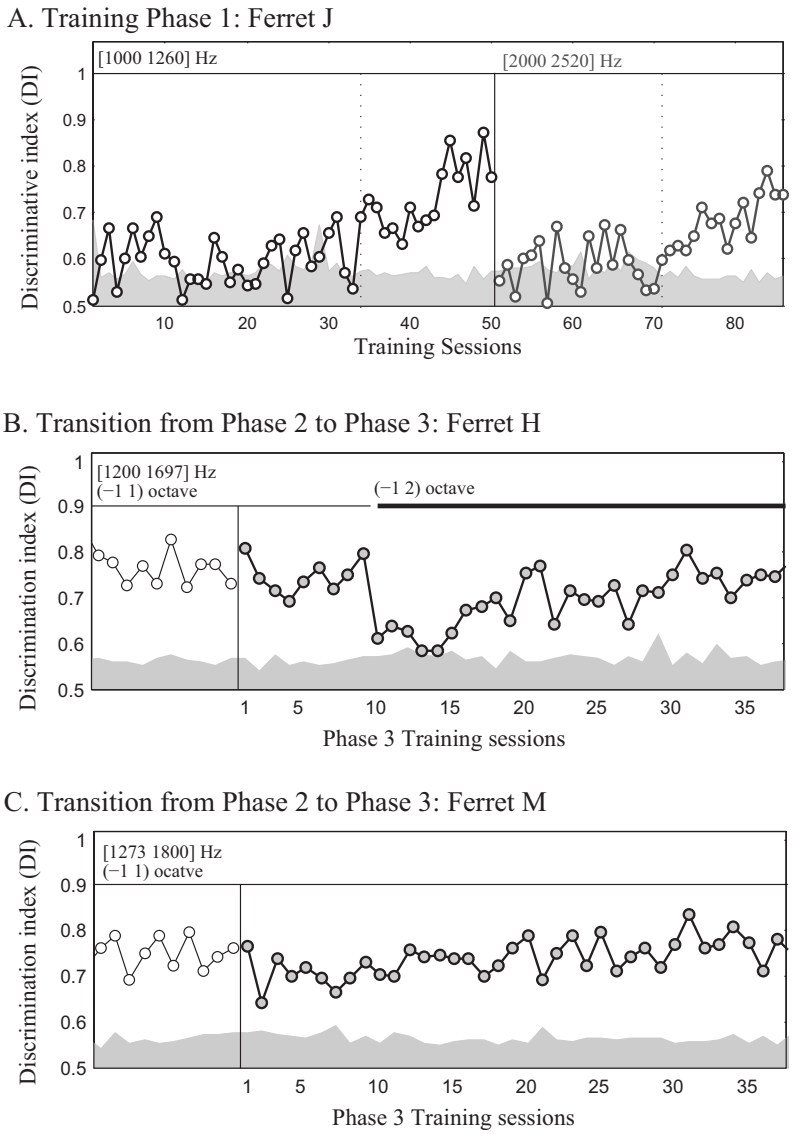


FIG. 5. *The frequency transposition in Phase-1 training and the transition from Phase 2 to Phase 3.* (A) The figure shows Phase-1 training data for ferret J. Phase-1 training was started (Training session=0) with a tone-pair at frequencies [1000 1260] Hz, and after learning the first pair, training was continued (Training session=50) with the second tone-pair at frequencies [2000 2520] Hz (*the vertical solid line*). [(B) and (C)] The animals maintained a high performance level when moved from Phase 2 (*open circles*) into Phase 3 (*filled circles*) when training within the same frequency range (indicated by the *thin horizontal line on the top of each figure*). In (B) the starting frequency range varied one octave above and below the initial frequency of the tone-pair [1200, 1697] Hz. Performance deteriorated when the frequency range of possible tone-pairs was extended an additional octave to two octaves above the initial tone-pair [indicated by the *thick horizontal line* in (B)] and regained after additional Phase-3 training. In (C) there was no change in frequency range during the transition from Phase 2 to Phase 3. The shaded area in (A)–(C) indicate the baseline performance (mean plus two standard deviation of the shuffled-DIs). A DI value above those dashed lines indicates a significant discriminative performance.

absolute frequencies. In order to test this conjecture, two ferrets were tested with a second tone-pair after learning the first tone-pair in Phase 1. The ferrets' behavioral discrimination initially plummeted with the new tone-pair, and the animals performed at chance level. The example shown in Fig. 5(A) is from ferret J. In Phase 1, ferret J was trained on its first tone-pair (1000 and 1260 Hz) for 34 sessions until it reached behavioral criterion (the first vertical dashed line). Performance initially declined to a random level [the shaded area in Fig. 5(A)] when a new tone-pair (2000 and 2520 Hz) was introduced in Training Session 50. It took another 21 training sessions (the second vertical dashed line) to reach behavioral criterion for consistent performance for the second tone-pair, though ferret J learned this discrimination faster than for the first tone-pair. This behavioral pattern was tested and replicated in ferret M. These behavioral results, showing that the animals did not generalize to new tone-pair in Phase 1, indicate that the animals' performance probably relied upon absolute pitch rather than relative pitch cues during Phase-1 training.

B. Transitioning from Phase 1 to Phase 2 and Phase 3: Learning relative pitch

Training in Phase 2 started after animals reached behavioral criterion on one (ferret H) or more tone-pairs (ferrets M and J) during Phase 1. In Phase-2 sessions, the reference and target stimuli in each trial consisted of upward and down-

ward versions of the same tone-pair (randomly chosen for each trial from a small set of tone-pairs) that remained constant throughout a given trial [see Fig. 2(B)]. The training began with a set of three tone-pairs near the frequencies of the last tone-pair in Phase-1 training. The number of tone-pairs used in Phase-2 training sessions was gradually expanded up to nine tone-pairs, spanning a two-octave frequency range [see the middle panel in Fig. 2(D)]. All three ferrets reached significant discriminative performance within the two-octave frequency range after 24 (ferret H), 68 (ferret M), or 77 (ferret J) training sessions. Since the absolute frequencies of tone-pairs in the reference and target sequence of a given trial changed on a trial-by-trial basis, using absolute cues (as in Phase 1) was no longer an efficient strategy for task performance. It is more likely that animals used relative pitch to solve the task than the alternative that the animals memorized the absolute cues and responses for each of the nine possible tone-pairs.

Two ferrets (H and M) progressed to Phase-3 training, in which each reference and target sequence in a trial was randomly chosen from a set of up to 17 tone-pairs which spanned four octaves [see Fig. 2(C) and the right panel in Fig. 2(D)]. Both animals maintained a significant discriminative performance when transitioning from Phase-2 to Phase-3 training, as shown in Figs. 5(B) and 5(C). These results indicate that the ferrets probably used the relative frequency contours of the sequence to solve the task in training Phase 2. However, even at this stage of training, animals did not generalize across all frequencies, and behavioral performance deteriorated when the frequency range of the tone-pairs expanded to a new frequency region, as indicated in Fig. 5(B). Apparently, the animals had generalized their performance only in a two-octave frequency range which had been achieved during Phase-2 training, and hence additional training was necessary to extend the discriminative performance to a larger frequency range.

Both ferrets (H and M) achieved significant discriminative performance with a four-octave frequency range after additional Phase-3 training, as illustrated in Fig. 4 (the light gray bars show Phase-3 performance for each ferret) and Fig. 6(A) (the two curves show performance over the frequency range for each ferret). Figure 6(A) shows the ferrets' discriminative performance for each of the tone-pairs after the animals reached performance criterion in Phase 3. In order to have sufficient behavioral data to compute the DI for each of the tone-pairs, the data in Fig. 6(A) were pooled from all ten sessions (the same data set as in Fig. 4). The significant discriminative performances were confirmed for all of the tone-pairs within the trained frequency range. The DIs from all tone-pairs within the four-octave frequency range [the lines with filled circle in Fig. 6(A)] were more than two standard deviations above the mean for shuffled-DIs [the dashed line in Fig. 6(A)]. The best performance was found at the tone-pairs in the middle frequency range for both ferrets. This result indicates that even though the ferrets could "partially" generalize the frequency contour categories within a wide frequency range after additional Phase-3 training, they were still subject to a frequency range constraint as in other nonhuman species (Hulse and Cynx, 1985; Izumi, 2001).

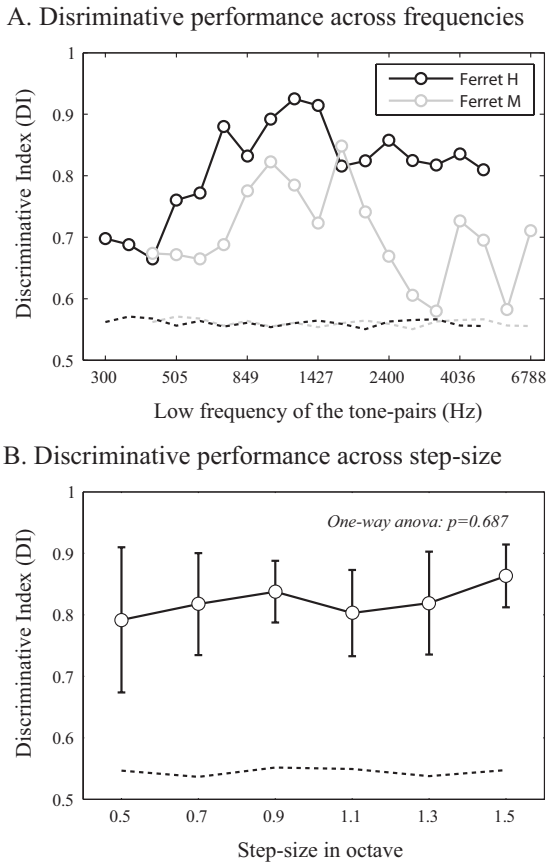


FIG. 6. *Discriminative performance across frequencies and step-sizes of the tone-pairs.* (A) Phase-3 data sets were the same as used in Fig. 4 for both ferrets M and H. The trials from all of those ten sessions were pooled together to compute the discriminative index for each of the tone-pairs. The significant discriminative performances were confirmed at all the tone-pairs within four-octave training frequency range. (B) Discriminative index at each frequency separation is represented as mean ± standard deviation (N =6). There is no significant difference in discriminative performance across the frequency separations between the component tones of the sequence. The horizontal dashed lines in (A) and (B) indicate the baseline performance (mean plus two standard deviation of the shuffled-DIs). A DI value above those dashed lines indicates a significant discriminative performance.

Those profiles of behavioral performance across the frequency range might be explained by the training history of the animals, which showed a similar profile of experienced frequencies.

C. The effect of frequency separation on task performance

In a separate set of experiments, the effects on task performance of the size of the frequency separation between the two component tones in the sequence were probed by a completely new set of two-tone sequences. As illustrated in Fig. 2(E), this stimulus set had novel two-tone sequences with variable frequency separation (0.5–1.5 octaves). Each reference and target was randomly picked from the stimulus list. One ferret was tested with this stimulus set alternately with the fixed standard frequency separation of 1/2 octave. There was no significant difference in behavioral performance across the different frequency separations [*one-way* ANOVA, p=0.687, Fig. 6(B)]. This result suggests that the animals did not attend to the interval size between the component tones

in the sequence, but simply to the direction of pitch change of the two-tone sequence. This result was also seen in human subjects in melody recognitions. The recognition of randomly generated melodies (novel melodies) was dominated by contours, while for familiar melodies both the contours and intervals were critical for recognition (Dowling and Fujitani, 1971; Dowling, 1978).

IV. GENERAL DISCUSSION

This study successfully demonstrated that ferrets are capable of relative pitch perception within a trained frequency range. With a three-phase training strategy, ferrets could learn to extract the relative pitch cue (the frequency contour) to discriminate between rising and falling two-tone sequences, independent of the absolute frequency of the tone sequence, over a four-octave range of frequencies.

During Phase-1 training, ferrets apparently used absolute pitch cues and discriminated target sequences from the reference sequences based on the initial or terminal pitch of the tone sequence. Transposition of the learned tone-pair to a different frequency caused a significant deterioration of performance, and animals needed to be retrained in order to master the new tone-pair. This result suggests that ferrets, like other nonhuman species, do not use relative pitch to solve contour discrimination tasks as a primary strategy, but are likely to use the absolute pitch of the tones in the sequence (particularly when the tone-pair frequencies are fixed).

However, ferrets were able to discriminate contour for multiple tone-pairs during Phase-2 training and appear to have learned to extract and utilize relative pitch cues in order to solve the task. Although the tone-pair frequencies remained constant within a given trial, and hence changes of the initial or ending pitch of the sequences could be used, ferrets appear to attend to the frequency contour within the sequence (relative cue) to solve the task. The behavioral evidence shown in Figs. 5(B) and 5(C) clearly indicates that performance on Phase 2 was very easily transferred to Phase 3, where the relative frequency contours were the only available cue. However, additional training was needed in Phase 3 in order to *expand* the frequency range over which the task was performed [indicated in Fig. 5(B)].

Sinnott *et al.* (1987) observed asymmetrical frequency discrimination in human subjects and some nonhuman primates and suggested that this asymmetrical sensitivity might relate to aspects of the species' vocal communication signal. However, in D'Amato's behavioral studies (reviewed in D'Amato, 1988), cebus monkeys and rats failed octave-generalization tests and also failed to demonstrate extraction of pitch contours. D'Amato concluded that perception of pitch contours requires specialized mechanisms that most animals lack and that "monkeys can't hum a tune … because they don't hear them" (D'Amato, 1988). This result was somewhat puzzling since birds and nonhuman primates are known to be capable of utilizing pitch cues and contours in vocal communication and recognition (Morton, 1977; Ratcliffe and Weisman, 1986; Weisman and Ratcliffe, 2004). In subsequent behavioral studies, rhesus monkeys were shown to generalize tonal melodies to whole one- and two-octave transpositions, but could not generalize over fractional transpositions, e.g., 0.5 octave or 1.5 octaves (Wright *et al.*, 2000). These observations are very intriguing, but also raise additional questions since adult human subjects can generalize for both tonal and atonal melodies over both octave and non-octave transpositions (McDermott and Hauser, 2005), suggesting that there may be limitations in relative pitch in nonhuman primates. There is also some behavioral evidence for perception and generalization of frequency contours to octave transpositions in one bottlenose dolphin (Ralston and Herman, 1995), but these claims need replication and further study.

The failure of ferrets to fully generalize to untrained frequency ranges in the present study suggests that ferrets were also subject to a frequency range constraint as described in other nonhuman species, including songbirds, rats, and monkeys (D'Amato and Salmon, 1984; Hulse and Cynx, 1985, 1986; Dooling *et al.*, 1987; Cynx, 1993; Wright *et al.*, 2000; Izumi, 2001). These nonhuman species have been found to have difficulty generalizing contour categories to novel frequency ranges and contours. Ferrets, however, did acquire good discriminative performance in the new frequency range in just a few training sessions [Fig. 5(B)]. This acquisition was much faster than the case with songbirds that required as many trials as they had in learning the original discrimination (Hulse and Cynx, 1985; Cynx, 1993). The frequency range constraint revealed the extent of absolute pitch perception in relative pitch perception. Although the relationship between the relative and absolute pitch perception remains unclear, the failure to acquire relative discrimination when eliminating the absolute cues suggests that relative pitch in songbirds may depend on first identifying the patterns on the basis of their absolute pitches (Page *et al.*, 1989). By contrast, in the present study, absolute pitch cues were available during only Phase-1 training, but not during the generalization training phases (Phases 2 and 3). Thus, animals learned the task with relative cues over a four-octave frequency range. These results indicate that ferrets, while not as good as humans, are somewhat better than songbirds in learning to utilize relative pitch cues.

Studies by Weisman *et al.* (1998) demonstrated that songbirds (even individuals reared in isolation) and parrots have highly accurate absolute pitch perception. In comparison, nonhuman mammals (such as rats) and humans exhibited only weak absolute pitch perception when classifying frequencies into ranges (Njegovan *et al.*, 1995; Weisman *et al.*, 1998, 2004). These findings lead Weisman to propose that there is a general difference in auditory processing in absolute and relative pitch perception between mammals (including humans) and songbirds (Weisman *et al.*, 2004). The results in ferrets are consistent with this hypothesis. Along with other recent behavioral studies on birds and monkeys (Page *et al.*, 1989; Wright *et al.*, 2000; Izumi, 2001; Brosch *et al.*, 2004), the present results provide evidence that animals can be trained to extract relative pitch when needed to perform a tonal pattern discrimination task.

V. CONCLUSIONS

The present study provides the evidence that ferrets can extract the tonal contour, independent of frequencies [Fig. 6(A)] and frequency separations [Fig. 6(B)] of the two-tone sequences. Appropriate task design and behavioral training procedures, such as the generalization training in Phase 2 in current study, are necessary to direct animal's attention to the relational features of the sequences and to develop a relational solution of the task. Similarly, an early study on birds (Page *et al.*, 1989) found that starlings extracted relative pitch from the pitch patterns only after acquiring a discrimination that permitted both absolute and relative pitch solutions. Although the natural tendency for animals may be to attend to the absolute properties of sound, the results of the present study suggest that they can still be trained to attend to relative pitch information.

ACKNOWLEDGMENT

This work was supported by R01DC005779 from NIDCD.

Bartlett, J. C., and Dowling, W. J. (**1980**). "Recognition of transposed melodies: A key distance effect in development perspective," J. Exp. Psychol. Hum. Percept. Perform. **6**, 501–515.

Brosch, M., Oshurkova, E., Bucks, C., and Scheich, H. (**2006**). "Influence of tone duration and intertone interval on the discrimination of frequency contours in a macaque monkey," Neurosci. Lett. **406**, 97–101.

Brosch, M., Selezneva, E., Bucks, C., and Scheich, H. (**2004**). "Macaque monkeys discriminate pitch relationships," Cognition **91**, 259–272.

Carterette, E. C., Friedman, M. P., and Cosmides, R. (**1965**). "Reaction time distributions in the detection of weak signal in noise," J. Acoust. Soc. Am. **38**, 531–542.

Cynx, J. (**1993**). "Auditory frequency generalization and a failure to find octave generalization in a songbird, The European starling (*Sturnus vulgaris*)," J. Comp. Psychol. **107**, 140–146.

D'Amato, M. R. (**1988**). "A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune," Music Percept. **5**, 453–480.

D'Amato, M. R., and Salmon, D. P. (**1984**). "Processing of complex auditory stimuli (tunes) by rats and monkeys (*Cebus apella*)," Anim. Learn Behav. **10**, 126–134.

Demany, L., and Armand, F. (**1984**). "The perceptual reality of tone chroma in early infancy," J. Acoust. Soc. Am. **76**, 57–66.

Dooling, R. J., Brown, S. D., Park, T. J., Okanoya, D., and Soli, S. D. (**1987**). "Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **101**, 139–149.

Dowling, W. J. (**1978**). "Scale and contour: Two components of a theory of memory for melodies," Psychol. Rev. **85**, 341–354.

Dowling, W. J., and Fujitani, D. A. (**1971**). "Contour, interval, and pitch recognition in memory for melodies," J. Acoust. Soc. Am. **49**, 524–531.

Emmerich, D. J., Gray, J., Watson, C., and Tanis, D. (**1972**). "Response latency confidence and ROCs in auditory signal detection," Percept. Psychophys. **11**, 65–72.

Fritz, J. B., Shamma, S., Elhilali, M., and Klein, D. (**2003**). "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex," Nat. Neurosci. **6**, 1216–1223.

Hulse, S. H., and Cynx, J. (**1985**). "Relative pitch perception is constrained by absolute pitch in songbirds (*Mimus, Molothrus and Sturnus*)," J. Comp. Psychol. **99**, 176–196.

Hulse, S. H., and Cynx, J. (**1986**). "Interval and contour in serial pitch perception by a passerine bird, The European starling (*Sturnus vulgaris*)," J. Comp. Psychol. **100**, 215–228.

Izumi, A. (**2003**). "Effect of temporal separation on tone-sequence discrimination in monkeys," Hear. Res. **175**, 75–81.

Izumi, A. (**2001**). "Relative perception in Japanese monkeys (*Macaca fuscata*)," J. Comp. Psychol. **115**, 127–131.

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, San Diego, CA).

McDermott, J. H., and Hauser, M. (**2005**). "The origins of music: Innateness, uniqueness, and evolution," Music Percept. **23**, 29–59.

Morton, E. S. (**1977**). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," Am. Nat. **111**, 855–869.

Njegovan, M., Ito, S., Mewhort, D., and Weisman, R. (**1995**). "Classification of frequencies into ranges by songbirds and humans," J. Exp. Psychol. Anim. Behav. Process **21**, 33–42.

Page, S. H., Hulse, S. H., and Cynx, J. (**1989**). "Relative pitch perception in the starling (*Sturnus vulgaris*): Further evidence for an elusive phenomenon," J. Exp. Psychol. Anim. Behav. Process **15**, 137–146.

Ralston, J. V., and Herman, L. M. (**1995**). "Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*)," J. Comp. Psychol. **109**, 268–277.

Ratcliffe, L., and Weisman, R. G. (**1986**). "Song sequence discrimination in the black-capped chickadee (*Parus atricapillus*)," J. Comp. Psychol. **100**, 361–367.

Sinnott, J. M., Owren, M. J., and Peterson, M. R. (**1987**). "Auditory frequency discrimination in primates: Species difference (*cercopithecus, macaca, homo*)," J. Comp. Psychol. **101**, 126–131.

Trehub, S. E., Bull, D., and Thorpe, L. A. (**1984**). "Infant's perception of melodies: The role of melodic contour," Child Dev. **55**, 821–830.

Trehub, S. E., and Hannon, E. E. (**2006**). "Infant music perception: Domain-general or domain-specific mechanisms?," Cognition **100**, 73–99.

Walker, K. M. M., Schnupp, J. W. H., Hart-Schnupp, S. M. B., King, A. J., and Bizley, J. K. (**2009**). "Pitch discrimination by ferrets for simple and complex sounds," J. Acoust. Soc. Am. **126**, 1321–1335.

Weisman, R., Njegovan, M., Sturdy, C., Phillmore, L., Cotle, J., and Mewhort, D. (**1998**). "Frequency-range discriminations: Special and general abilities in zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*)," J. Comp. Psychol. **112**, 244–258.

Weisman, R. G., Njegovan, M. G., Williams, M. T., Cohen, J. S., and Sturdy, C. B. (**2004**). "A behavior analysis of absolute pitch: Sex, experience, and species," Behav. Processes **66**, 289–307.

Weisman, R. G., and Ratcliffe, L. (**2004**). "Relative pitch and the song of black-capped chickadees," Am. Sci. **92**, 532–539.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., and Neiworth, J. J. (**2000**). "Music perception and octave generalization in rhesus monkeys," J. Exp. Psychol. Gen. **129**, 291–307.

Sound emission and reception tuning in three cicada species sharing the same habitat

Jérôme Sueur[a)]
Muséum National d'Histoire Naturelle, Département Systématique et Evolution, UMR 7205 CNRS OSEB, 45 rue Buffon, 75005 Paris, France

James F. C. Windmill
Centre for Ultrasonic Engineering, Department of Electronic and Electrical Engineering, University of Strathclyde, Royal College Building, 204 George Street, Glasgow G1 1XW, United Kingdom

Daniel Robert
University of Bristol, School of Biological Sciences, Woodland Road, Bristol BS8 1UG, United Kingdom

(Received 1 April 2009; revised 18 December 2009; accepted 18 December 2009)

Many animal species acoustically communicate at the same place and time generating complex acoustic environments. However, the acoustic parameter space is usually structured, with each species emitting identifiable signals. While signal partitioning has been reported, very few analyses include the mechanical spectral response of auditory organs. The loud chorus generated by three cicadas (*Cicada orni*, *Cicadatra atra*, and *Lyristes plebejus*) was studied. The vibration pattern of *L. plebejus* shows traveling waves as previously observed in *Ctra. atra*. The spectral properties of both calling songs and tympanal auditory systems primarily indicate that each species uses its own frequency band. Male tympanal membranes (TMs) are tuned to their own song's dominant frequency, except for *C. orni*, which is sensitive to the lowest frequency band of its song. In contrast, female TMs are broadly tuned to the male songs. *Ctra. atra* females differ by tuning to frequencies slightly higher than the male song. Hence, acoustic space partitioning occurs for both emitter and receiver, but does not seem to fully preclude interference risk as some spectral overlap exists. In addition to the local physical ecology of each species, selective attention to conspecific signals is likely to be enhanced by further mechanical and neuronal processing.
 [DOI: 10.1121/1.3291036]

PACS number(s): 43.80.Lb, 43.80.Ka, 43.80.Jz [MJO] Pages: 1681–1688

I. INTRODUCTION

Animals communicating with sound are often simultaneously active in the same habitat. They generate multispecies choruses with intense acoustic interference, each calling species being a source of noise to the others (Römer, 1993). Several strategies to increase the information transfer have been hypothesized (Brumm and Slabbekoorn, 2005). One strategy is the production of species-specific signals, leading to a partitioning of the acoustic parameter space through temporal and/or frequency exclusion in accordance with the signal space concept and the matched-filter hypothesis (Capranica and Moffat, 1983; Nelson and Marler, 1990). Partitioning has been particularly illustrated in frogs, crickets, and cicadas (Hödl, 1977; Riede, 1993; Lüddecke *et al.*, 2000; Sueur, 2002; Diwakar and Balakrishnan, 2007a). The validity of a signal partitioning strategy is linked, however, to the sensory capacity of the receivers, in this case the spectral sensitivity of the auditory system.

In this study, the particular role of peripheral auditory processing in the detection and selection of species-specific signals through frequency selection was investigated for three cicada species, *Cicada orni*, *Cicadatra atra*, and *Lyristes plebejus*. The male songs of these three cicada species are broadcasted within the same habitat and at the same time (Boulard and Mondon, 1995). Hundreds of specimens can be found on a single tree generating a continuous and loud chorus (Fig. 1, electronic supplementary material sound S1). Using scanning laser Doppler vibrometry on field-collected animals, the vibrations of *L. plebejus* tympanal membrane (TM) are first described and then compared with those of *C. orni* and *Ctra. atra*, which have previously revealed traveling and standing waves, respectively (Sueur *et al.*, 2006, 2008a, 2008b). The frequency characteristics of both the calling song (CS), together with the mechanical response of the tympanal membrane, are then quantified and compared across the three species. The study reveals that *L. plebejus* TM moves following a traveling wave pattern. Furthermore, these three closely related cicadas share acoustic space through frequency exclusion. Finally, it is shown that some frequency selectivity takes place at the peripheral level of the TM, although this tuning is quite broad and does not completely match the CS dominant frequency.

II. METHOD

A. Calling song recording and analysis

The CS of 15 males per species (*C. orni*, *Ctra. atra*, and *L. plebejus*) were recorded in France from 1995 to 2008 and

a)Author to whom correspondence should be addressed. Electronic mail: sueur@mnhn.fr

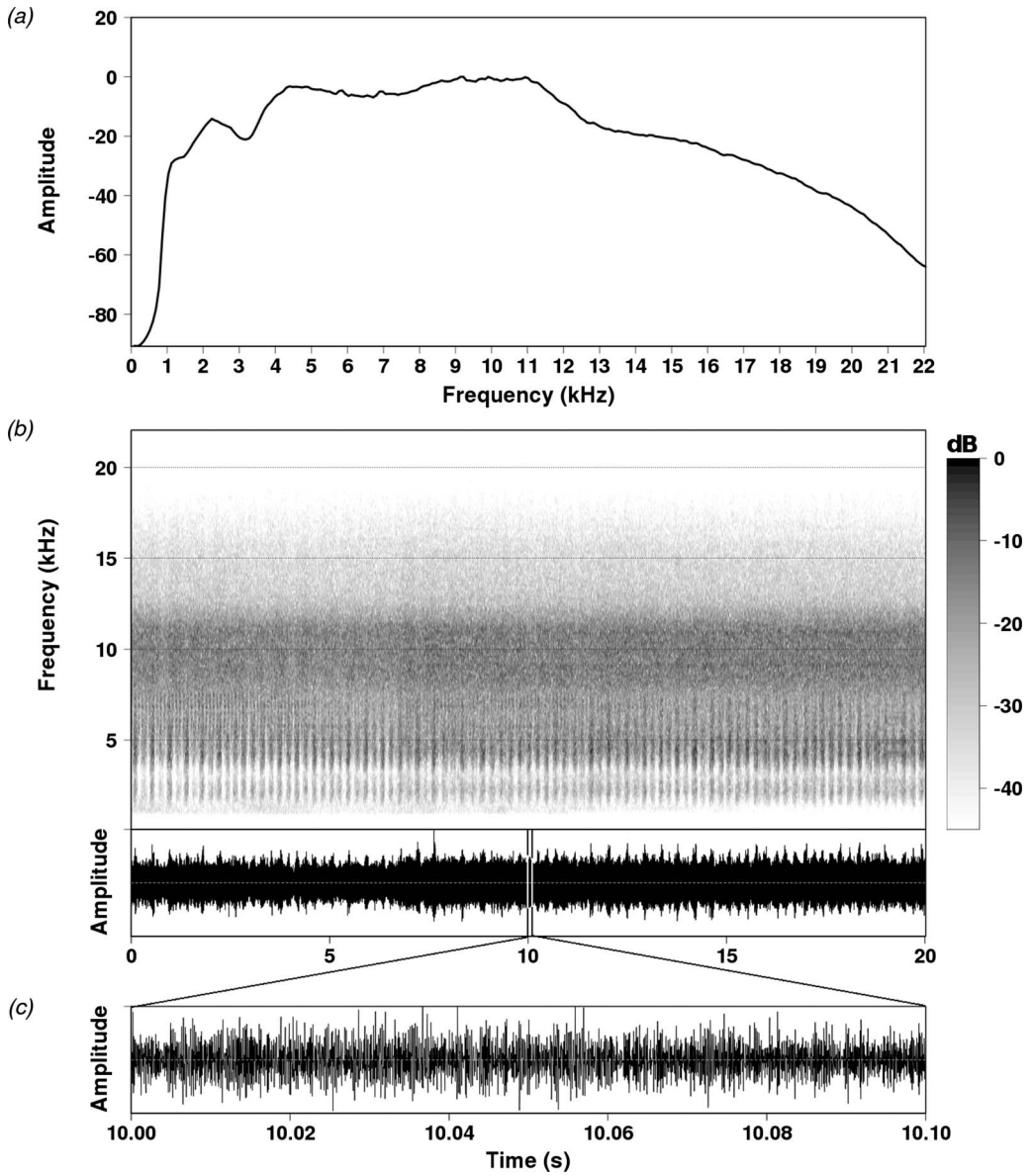


FIG. 1. Chorus of *C. orni*, *Ctra. Atra*, and *L. plejus* recorded in Cuges-les-Pins, France. (a) Mean spectrum, (b) spectrogram and oscillogram, and (c) detailed oscillogram corresponding to the zone delimited by the vertical lines on the oscillogram depicted in (b). A 1-kHz high-pass filter was applied to the sound to remove background noise due to wind. Short-term Fourier analysis parameters: Hamming window, frequency resolution=12.5 Hz, and dynamic range =45 dB.

analyzed in the frequency domain. Ten recordings of *C. orni* made in 2001 were already used in Sueur *et al.*, 2006, and four recordings of *Ctra. atra* made in 2006 in Sales-le-Château were already used in Sueur *et al.*, 2006. Sampling details for each species can be found in Table I. Isolated males were recorded from a distance of 0.5–1 m using a Telinga Pro4PiP microphone connected to a Sony TCD-D8 digital recorder or with a Sennheiser ME64 microphone connected to a Marantz PMD 670 digital recorder. The sampling frequency was 44.1 kHz. 1 min of each CS was analyzed in the frequency domain using the package SEEWAVE (Sueur *et al.*, 2008a), which is implemented in the R environment (R Development Core Team, 2009). A mean frequency spectrum was computed using a sliding Fourier transform (12.5 Hz resolution, nonoverlapping Hamming window).

B. Tympanal membrane mechanics

Specimens were caught and studied in 2005, 2007, and 2008. Specimens of *Ctra. atra* collected in 2005 (15 specimens) were used in Sueur *et al.*, 2008a, 2008b but mean spectra data were not reported. Data for *C. orni* collected in 2007 (19 specimens) were extracted from Sueur *et al.*, 2008a, 2008b and completed by new data obtained in 2008. Sampling details for each species can be found in Table I. Animals were cooled to 8–10 °C, transferred to Bristol (UK) in an icebox, and then placed at 24–26 °C and 40%–62% relative humidity for >2 h before measurements. The sound-induced vibrations of the TM were measured with a Polytec PSV-300-F microscanning laser Doppler vibrometer, with an OFV-056 scanning head. One ear was examined per animal. The wings and the legs were cut back before me-

TABLE I. Location of recordings and specimens collected for TM mechanics. Recordings of *C. orni* made in 2001 in Peyriac-de-Mer and Molitg-les-Bains were already used in Sueur *et al.* (2008b). Recordings of *Ctra. atra* made in 2005 in Sales-le-Château in 2005 were already used in Sueur *et al.* (2006). Specimens of *C. orni* and *Ctra. atra* and collected for TM mechanics in 2005 and 2007 were previously used in Sueur *et al.* (2006, 2008b), respectively.

Locality	Administrative division	Date	Sample size
	Calling song recordings		
	C. orni		
Peyriac-de-Mer	Aude (11)	07-16-2001	5
Molitg-les-Bains	Pyrénées-Orientales (66)	07-17-2001	5
Saint-Hypolyte	Pyrénées-Orientales (66)	07-17-2001	5
	Ctra. atra		
Lafare	Vaucluse (84)	06-30-1996	3
Lespignan	Aude (11)	06-30-2000	1
Sales-le-Château	Pyrénées-Orientales (66)	06-22-2005	4
Cuges-les-Pins	Bouches-du-Rhône (13)	07-06-2008	7
	L. plebejus		
Beaumes-de-Venise	Vaucluse (84)	07-11-1995	3
Beaumes-de-Venise	Vaucluse (84)	07-13-1995	1
Cairanne	Vaucluse (84)	07-13-1995	1
Cuges-les-Pins	Bouches-du-Rhône (13)	07-06-2008	5
Riboux	Bouches-du-Rhône (13)	07-06-2008	2
Castellet	Bouches-du-Rhône (13)	07-07-2008	1
Signes	Bouches-du-Rhône (13)	07-07-2008	1
Gémenos	Bouches-du-Rhône (13)	07-07-2008	1
	TM mechanics		
	C. orni		
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2007	19
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2008	5
	Ctra. atra		
Sales-le-Château	Pyrénées-Orientales (66)	06-22-2005	15
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2007	3
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2008	8
	L. plebejus		
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2007	2
Cuges-les-Pins	Bouches-du-Rhône (13)	07-09-2008	4
Gémenos	Bouches-du-Rhône (13)	07-09-2008	13

chanical measurements. The operculum and the meracanthus that ventrally cover the tympanal organs, but are not mechanically linked to them, were also removed. Animals were not anaesthetized during measurements but were firmly attached to a horizontal brass bar (6 mm wide, 1 mm thick, and 16 mm long) using BLU-TACK (Bostik-Findley, Stafford, UK). The overall spectral properties of the TM were assessed by orienting the animals such that the vibrometer could scan the TM as completely as possible, perpendicular to its surface. The tympanal ridge (TR) area closest to the apodeme could not always be examined, as in *C. orni* and *L. plebejus* females where it was hidden by a cuticular sternal expansion that could not be removed without compromising the auditory system's mechanical integrity. The mechanical deflections of *L. plebejus* tympanum were characterized in detail to ensure a comparison with those of *C. orni* and *Ctra. atra* that were previously described (Sueur *et al.*, 2006, 2008b).

Experiments were carried out on a TMC 784-443-12R vibration isolation table located in a 20 m^3 acoustic isolation booth (Industrial Acoustics IAC series 1204A). TM vibrations were examined in response to frequency modulated (FM) signals (80-ms duration) sweeping at similar intensity all frequencies from 1 to 22.05 kHz (±1 dB). Acoustic stimuli were amplified with a TAFE570 Sony Amplifier, and broadcasted 0.25 m from the cicada with an ESS AMT-1 loudspeaker. Stimulation intensity was 66 dB sound pressure level (SPL) at the cicada. Reference SPL was measured using a Bruel & Kjaer 4138 microphone and Bruel & Kjaer 2633 preamplifier, with a linear response in the measured frequency range. Microphone sensitivity was calibrated using a Bruel & Kjaer 4231 sound level calibrator. The microphone was positioned 10 mm from the tympanum, with its diaphragm parallel to the direction of sound propagation. Signals were simultaneously sampled at 102.4 kHz. Sets of 15 data windows were acquired and averaged for each measurement point. Using a fast Fourier transform (rectangular window), a frequency spectrum was produced for the microphone and laser signals with a resolution of 12.5 Hz. Vibrometry frequency spectra were normalized to those of the reference microphone by the computation of the following transfer function:

$$H(f) = \frac{G_{ab}(f)}{G_{aa}(f)}, \tag{1}$$

where $G_{ab}(f)$ is the cross-spectrum of the reference and velocity signals and $G_{aa}(f)$ is the autospectrum of the reference signal. The data reliability was estimated by means of magnitude-squared coherence (C) functions between the laser and microphone signals, as follows:

$$C(f) = \frac{G_{ab}(f)G_{ba}(f)}{G_{aa}(f)G_{bb}(f)}, \tag{2}$$

where $G_{ab}(f)$ is the cross-spectrum of the velocity signal and reference signal (Fourier transform of the cross correlation function between the two time series), $G_{ba}(f)$ is the cross-spectrum of the reference signal and velocity signal, $G_{aa}(f)$ and $G_{bb}(f)$ are, respectively, the autospectra of the reference signal and of the velocity signal. Only highly coherent measurements (>85%) with minimal contamination by unrelated noise were used.

C. Simulated chorus spectrum and tympanal spectral filters

To test the filtering effect of the TM on the sound produced by a chorus, the CS mean spectrum of each species was normalized and summed. This spectrum was multiplied by the normalized TM mean spectrum of each species and of each sex. A sound file was built adding at equal amplitude level from each of the three species (electronic supplementary material sound S2). Simulated chorus finite impulse response (FIR) filters were applied to this summed sound file (Stoddard, 1998). To apply this filter, an inverse Fourier transform of TM spectral properties of each species and sex

was computed. This transform resulted in a series of impulse responses. Each of them was convolved with the summed sound file, as follows:

$$y(t) = (h * x)(t), \tag{3}$$

with $y(t)$ the filtered signal, $x(t)$ the original signal, and $h(t)$ the impulse response.

D. Spectral description and comparison

For each spectrum, the dominant frequency (frequency peak of highest energy) and the spectral flatness measure (SFM) were estimated. SFM was calculated as follows:

$$\mathrm{SFM} = n \times \frac{\sqrt[n]{\prod f_i}}{\sum f_i}, \tag{4}$$

where n is the number of f_i frequency bands constituting the spectrum. Values of SFM lie between 0 and 1, which, respectively, indicate a highly tuned resonating system and a nonresonating system.

Overlap between the spectra was assessed with the index (I) as follows:

$$I = 1 - \frac{\sum (S_1(f_i) - S_2(f_i))}{2}, \tag{5}$$

where $S_1(f)$ and $S_2(f)$ are the probability mass functions of the spectra to be compared.

Values of I range from 0 to 1, respectively, indicating nonoverlapping and similar spectra. Originally expressed as amplitude data (mV for recorded songs) or gain data (nm Pa^{-1} for TM vibrations), spectra were normalized to a maximum of 1 before computing mean spectra and estimating interval confidence. All statistics were computed using R (R Development Core Team, 2009).

III. RESULTS

A. Vibration pattern of *L. plebejus* tympanum

The tympanum of *L. plebejus* showed the same general structure as found in *C. orni* and *Ctra. atra* with a thin transparent membrane crossed by a spearlike TR (Sueur *et al.*, 2006, 2008a, 2008b). As in the two other species, there was a clear sexual dimorphism in *L. plebejus* affecting both size and shape (Fig. 2). The male tympanum was large and rectangular, the TR being small compared to the whole surface of the membrane. The female tympanum was smaller, had a triangular shape, and showed a dark and large TR whose length was almost three-quarters that of the largest membrane width. When scanning the whole surface of the male TM, it appeared that the central zone moved much less than the lateral ones. In particular, the lateral area where the TR was located showed a typical traveling wave going along the TR from its apex to its base.

The envelope of the wave in motion along the TR was particularly apparent when the response was displayed every 40° of phase angle in the full oscillation cycle (Fig. 3). This motion was confirmed by analyzing the phase lag as a function of stimulus frequency. Phase lag was substantial, exceeding the phase lag expected for the displacement response of a simple oscillator (maximum of −90° at frequencies above resonance). Lag increased as a function of frequency. At higher frequencies, such as 15 kHz, the phase lag reached −600° (Fig. 4). When stimulated, the whole female tympanum moved and the TR was similarly crossed by a traveling wave going from the apex to the base. The displacement amplitude of the traveling wave was ten times smaller than the one observed in males (Fig. 3). The phase lag reached −150° at 15 kHz (Fig. 4).

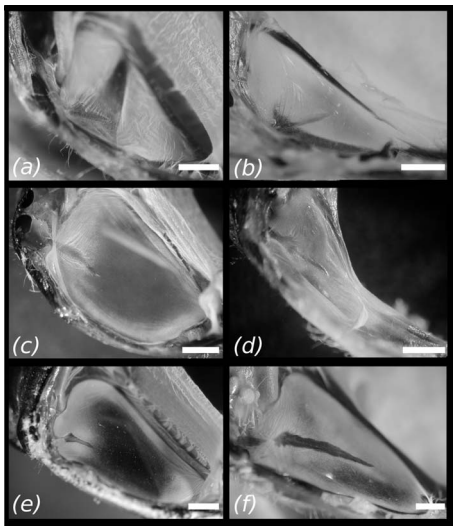


FIG. 2. Posterior view of female (left column) and male (right column) tympanum of [(a) and (b)] *C. orni*, [(c) and (d)] *Ctra. Atra*, and [(e) and (f)] *L. plebejus*. Note the different scales. Scale bars=0.5 mm.

B. Calling songs of the three species

Each species' CS occupied a specific frequency channel minimizing acoustic competition. CS spectra of the three

FIG. 3. Envelopes of mechanical deflections along *L. plebejus* TR from its apex (x) to its base (y). Deflections are shown for phase increments of 40° in the full oscillation cycle. Top: male TR driven at 5.475 kHz. Bottom: female TR driven at 6.925 kHz. In each case, the driving frequency is the frequency of sound played back that deflected the TR the most, i.e., that induced the mean greatest deflection. Note the change of y left scale.

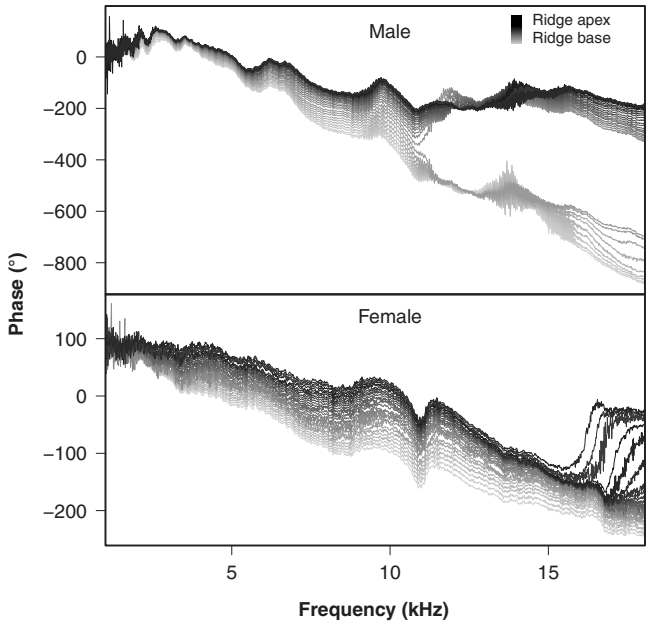


FIG. 4. Phase response along *L. plebejus* TR as a function of the driving frequency. The phase response shows an increasing lag with both frequency and distance along the TR. The phase lag increases to −245° for females and to −880° for males.

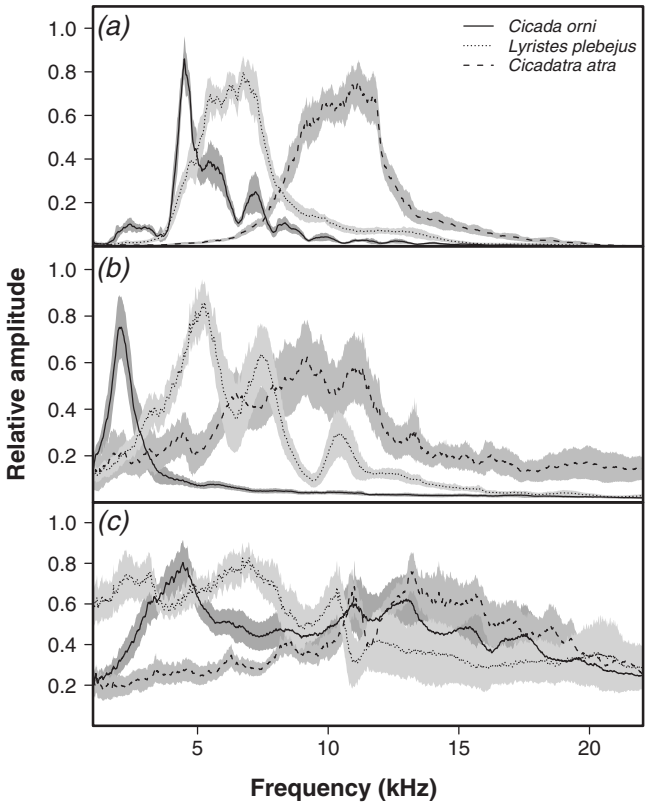


FIG. 5. Frequency spectra of the CS, and of both sex's TM vibrations. (a) Calling song, (b) male TM, and (c) female TM. Bold lines (*C. orni*: solid, *Ctra. atra*: dotted, and *L. plebejus*: dashed) refer to the mean of the individual spectra and surface to the 95% confidence interval. Spectra were normalized to a maximum of 1 before mean and interval confidence estimations. Sample sizes: 15 *C. orni*, 15 *Ctra. atra*, and 15. *L. plebejus* for CS; 13 *C. orni*, 13 *Ctra. atra*, and 10 *L. plebejus* for male TM; 11 *C. orni*, 13 *Ctra. atra*, and 9 *L. plebejus* for female TM.

species showed similar low flatness values but different dominant frequencies [see Fig. 5(a) and Table II]. CS spectra did not overlap except for *C. orni* and *L. plebejus* between 4 and 6 kHz, leading to a higher similarity index than between *Ctra. atra* and *C. orni*, and *Ctra. atra* and *L. plebejus*, respectively (Table III).

C. Tympanal spectral characteristics of the three species

Male TMs of *C. orni* and *L. plebejus* were sharply tuned, and the male TM of *C. orni* was not tuned to its own CS. Male TM spectra showed different flatness and dominant frequency values [Fig. 5(b), Table II]. The TM dominant frequency corresponded to the CS dominant frequency, except for *C. orni* where an important mismatch appeared (Welch *t*-test for the dominant frequency, $df=1$, $p<0.001$). The similarity index indicated the highest values occurring between male TM and CS spectra of the same species (Table III).

Female TMs of *C. orni* and *L. plebejus* were broadly tuned, and female TMs of *Ctra. atra* were not sharply tuned to their conspecific CS. Female TM showed similar high flatness values, but different dominant frequencies [Fig. 5(c), Table II]. The dominant frequency of the female TM of *C. orni* and *L. plebejus* matched the dominant frequency of their CS, whereas the *Ctra. atra* female TM dominant frequency was significantly higher (Welch *t*-test, $df=1$, $p<.001$). The similarity index showed the highest values between the female TM and CS spectra of the same species except in the case of *C. orni*, for which an additional sensitivity was present at higher frequencies (Table III).

The TMs of *C. orni* males and females acted as sharp and broad low-pass filters, respectively [Fig. 6(a); electronic supplementary material sounds S3 and S4]. The TMs of *Ctra. atra* acted as broad high-pass filters [Fig. 6(b); electronic supplementary material sounds S5 and S6]. The TMs of *L. plebejus* acted as broad low-pass filters [Fig. 6(c); electronic supplementary material sounds S7 and S8].

IV. DISCUSSION

The chorus of *C. orni*, *Ctra. atra*, and *L. plebejus* generates a complex acoustic environment. However, each species shows a specific calling spectrum in the frequency domain, leading to a partitioning in accordance with the signal space concept and the matched-filter hypothesis (Capranica and Moffat, 1983; Nelson and Marler, 1990). In this context, this work shows that the TM not only operates as an acousticomechanical transducer, but also as a frequency filter that contributes to signal detection and recognition. The TMs of the three species, except for the males of *C. orni*, show resonance peaks that match the highest energy peak of the conspecific CS. The TM detuning of male *C. orni* does not produce a total filtering of the surrounding noise as the filter frequency response is not totally flat above 3 kHz. However, it significantly increases the lower frequency band of the chorus, thus selectively enhancing the component produced by *C. orni* and reducing signal competition with *L. plebejus* in the 4–6 kHz range. This also prevents the male from being deafened by its own song (Sueur *et al.*, 2008b).

TABLE II. Spectral properties of the CS and the TM. Values of spectral flatness could range from 0 to 1, which, respectively, are indicative of a highly tuned resonating system and a nonresonating system. Mean values are compared with an analysis of variance.

Spectral parameters	*C. orni*	*Ctra. atra*	*L. plebejus*	$F_{d.f}$	*p*
		CS			
Dominant frequency (kHz)	4.51 ± 0.15 (*n*=15)	10.78 ± 0.86 (*n*=15)	6.57 ± 0.67 (*n*=15)	$376.4_{2,42}$	<0.001
Spectral flatness	0.21 ± 0.04 (*n*=15)	0.27 ± 0.08 (*n*=15)	0.22 ± 0.09 (*n*=15)	$2.1_{2,42}$	0.135
		Male TM			
Dominant frequency (kHz)	2.13 ± 0.28 (*n*=13)	10.23 ± 1.53 (*n*=13)	5.48 ± 0.77 (*n*=10)	$207.1_{2,33}$	<0.001
Spectral flatness	0.56 ± 0.08 (*n*=13)	0.76 ± 0.08 (*n*=13)	0.55 ± 0.04 (*n*=10)	$39.58_{2,33}$	<0.001
		Female TM			
Dominant frequency (kHz)	4.39 ± 0.92 (*n*=11)	13.29 ± 1.76 (*n*=13)	6.93 ± 1.02 (*n*=9)	$140.6_{2,30}$	<0.001
Spectral flatness	0.92 ± 0.02 (*n*=11)	0.90 ± 0.05 (*n*=13)	0.88 ± 0.06 (*n*=9)	$2.11_{2,30}$	0.139

The mechanical aspect of peripheral signal processing found in the cicada is similar to that already documented for several grasshopper and moth species (Windmill *et al.*, 2006, 2008). In particular, the TM deflection pattern of *L. plebejus* measured in this work shows a convergence with the TM vibrations previously described for *Ctra. atra* (Sueur *et al.*, 2006) with a traveling wave that goes along the TR. *C. orni* appears then as unique since the TM moves following a standing wavelike behavior with a simple up-and-down oscillation (Sueur *et al.*, 2008b). The origin of this notable difference in modes of vibrations has still to be discovered but is probably linked to differences in membrane shape, thickness, and stiffness.

Each different cicada TM does not filter out all of the background noise due to other cicada species. This is particularly true for females, which show tympanal mechanical responses with broad spectra. To respond appropriately to the conspecific male CS and no other, a female must decode the species-specific information embedded within the signal alone or in a chorus. Several mechanisms very likely contribute to this process, starting with the reception of sound by the TM (Robert and Hoy, 2007) and interneuronal selective responses (Fonseca *et al.*, 2000), altogether supporting the recognition of conspecific CS. In the present experiments, the TR apex was not always optically accessible. Hence the resonant frequency peak measured may be underestimated, in particular, for *L. plebejus* and *C. orni* females. The apex resonance may be lower and sharper. Connected to the auditory neurons by an apodeme, the TR is a key structure that, through its vibration, transmits frequency information to the nervous system. The exact mechanisms of information transfer between TR and sensory neurons are unknown but the apodeme of *C. atra* has been recently proved to work as an additional frequency filter that sharpens frequency tuning around the CS dominant frequency (Windmill *et al.*, 2009). There is also some neurophysiological evidence for another filter enhancing frequency selection, as fine frequency analysis can be achieved by auditory interneurons (Fonseca *et al.*, 2000).

Evidently, in addition to spectral composition, the temporal structure of the CS is also important for species-specific recognition (Gerhardt and Huber, 2002). Within the chorus, to the human ear, species-specific patterns seem to be lost. Yet, previous research has shown that temporal domain selective audition is possible through the action of neuronal temporal filters, as illustrated in cicadas (Huber *et al.*, 1990) and katydids (Schul, 1998). Another way to reduce interference is the use of substrate-borne signals that could be trans-

TABLE III. Similarity (*I*) between the CS and TM spectra. Values of *I* could range from 0 to 1, which, respectively, are indicative of nonoverlapping and similar spectra. Bold style indicates for each species the correspondence between CS spectrum and TM spectrum (table diagonal).

	CS spectrum		
	C. orni	*L. plebejus*	*Ctra. atra*
	CS spectrum		
C. orni	0.76 ± 0.06 (*n*=105)		
L. plebejus	0.60 ± 0.08 (*n*=195)	0.83 ± 0.05 (*n*=105)	
Ctra. atra	0.20 ± 0.06 (*n*=195)	0.30 ± 0.05 (*n*=195)	0.82 ± 0.06 (*n*=105)
	TM spectrum		
C. orni (male)	**0.35 ± 0.04 (*n*=195)**	0.32 ± 0.05 (*n*=195)	0.25 ± 0.04 (*n*=195)
C. orni (female)	**0.43 ± 0.06 (*n*=165)**	0.45 ± 0.06 (*n*=165)	0.45 ± 0.07 (*n*=165)
L. plebejus (male)	0.65 ± 0.05 (*n*=150)	**0.65 ± 0.06 (*n*=150)**	0.34 ± 0.05 (*n*=150)
L. plebejus (female)	0.48 ± 0.06 (*n*=135)	**0.51 ± 0.06 (*n*=135)**	0.41 ± 0.06 (*n*=135)
Ctra. atra (male)	0.40 ± 0.10 (*n*=195)	0.47 ± 0.08 (*n*=195)	**0.56 ± 0.09 (*n*=195)**
Ctra. atra (female)	0.31 ± 0.05 (*n*=195)	0.35 ± 0.05 (*n*=195)	**0.46 ± 0.07 (*n*=195)**

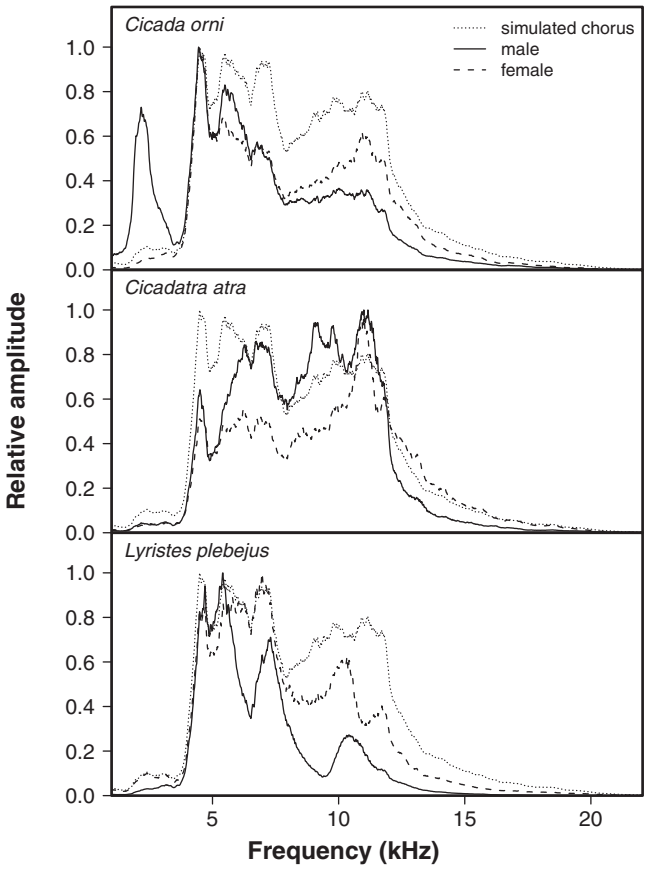


FIG. 6. Frequency-filter effect of the TM on a simulated chorus spectrum. Dotted gray lines refer to the summed spectrum of the CS mean spectra of the three species, and black lines refer to the summed spectrum when filtered by the TM (male: solid; female: dashed). All spectra were normalized to a maximum of 1.

mitted at close range when partners stay on the same plant part (Cocroft and Rodríguez, 2005) and the selection by each species of a particular microhabitat (Hödl, 1977; Sueur, 2002; Diwakar and Balakrishnan, 2007b). Here each species seems to occupy a different position in the tree: *L. plebejus* at the tip of canopy thin branches, *C. orni* on trunk and large branches as already reported (Claridge *et al.*, 1979), and *Ctra. atra* on small branches.

V. CONCLUSIONS

This study reveals that three closely related cicadas share acoustic space through frequency exclusion. It is also shown that some frequency selectivity takes place at the peripheral level of the TM. However, this peripheral tuning is quite broad, not perfectly matching the CS dominant frequency. Comparing the three species considered here, it is apparent that the peripheral processing achieved by the TM has evolved to carry out at least two tasks: the transduction of acoustic energy and the early processing of sound-induced vibrations. It should be noted that the actual frequency selectivity of the different TMs may be higher than reported here, as it was not always possible to measure the entire TM area with the laser vibrometry technique. It is also unclear how the peripheral (mechanical) frequency selectivity of the TM relates to the further stages of auditory processing, with the transfer of information to the sensory neurons, the interneurons, up to the brain. Finally, this work describes the TM response of only three, closely related, cicadas, yet has shown startling differences in the mechanics of those responses between one species and the others. There are approximately 2500 species of cicada worldwide, suggesting that there may be many further different peripheral auditory mechanisms still to be discovered.

ACKNOWLEDGMENTS

This work was financially supported by a Co. of Biologists JEB Travel Fellowship and by a grant from the British Council in France. We are indebted to Stéphane Puissant and Sofia Seabra for their assistance when recording *C. orni*. D.R. is a research fellow of the Royal Society London.

Boulard, M., and Mondon, B. (**1995**). *Vies et Mémoires de Cigales (Life Histories and Memories of Cicadas)* (L'imagier, Equinoxe, Barbentane).
Brumm, H., and Slabbekoorn, H. (**2005**). "Acoustic communication in noise," Adv. Study Behav. **35**, 151–209.
Capranica, R. R., and Moffat, J. M. (**1983**). "Neurobehavioral correlates of sound communication in Anurans," in *Advances in Vertebrate Neuroethololgy*, edited by J.-P. Ewert, R. R. Capranica, and D. J. Ingle (Plenum Press, New York), pp. 701–730.
Claridge, M. F., Wilson, M. R., and Singhrao, J. S. (**1979**). "The songs and calling sites of two European cicadas," Ecol. Entomol. **4**, 225–229.
Cocroft, R. B., and Rodríguez, J. G. (**2005**). "The behavioral ecology of insect vibrational communication," Bioscience **55**, 323–334.
Diwakar, S., and Balakrishnan, R. (**2007a**). "The assemblage of acoustically communicating crickets of a tropical evergreen forest in southern India: Call diversity and diel calling pattern," Bioacoustics **16**, 113–135.
Diwakar, S., and Balakrishnan, R. (**2007b**). "Vertical stratification in an acoustically communication Ensiferan assemblage of a tropical evergreen forest in southern India," J. Trop. Ecol. **23**, 479–486.
Fonseca, P. J., Münch, D., and Hennig, R. M. (**2000**). "How cicadas interpret acoustic signals," Nature (London) **405**, 297–298.
Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans* (University of Chicago Press, Chicago).
Hödl, W. (**1977**). "Call differences and calling site segregation in Anuran species from Central Amazonian floating meadows," Oecologia **28**, 351–363.
Huber, F., Kleidienst, H., Moore, T. E., Schildberger, K., and Weber, T. (**1990**). "Acoustic communication in periodical cicadas: Neuronal responses to songs of sympatric species," in *Sensory Systems and Communication in Arthropods*, edited by F. G. Gribakin, K. Wiese, and A. V.Popov (Birkhaüser, Basel), pp. 217–228.
Lüddecke, H., Amézquita, A., Bernal, X., and Guzmán, F. (**2000**). "Partitioning of vocal activity in a neotropical highland-frog community," Stud. Neotrop. Fauna Environ. **35**, 185–194.
Nelson, D. A., and Marler, P. (**1990**). "The perception of birdsong and an ecological concept of signal space," in *Comparative Perception: Complex Signals*, edited by M. Berkley and W. Stebbins (Wiley, New York), pp. 443–478.
R Development Core Team (**2009**). "R: A language and environment for statistical computing," Vienna, Austria, http://www.R-project.org (Last viewed 12/18/09).
Riede, K. (**1993**). "Monitoring biodiversity: Analysis of Amazonian rainforest sounds," Ambio **22**, 546–548.
Robert, D., and Hoy, R. R. (**2007**). "Auditory systems in insects," in *Invertebrate Neurobiology*, edited by R. Greenspan and G. North (Cold Spring Harbour Laboratory, San Diego, CA), pp. 155–184.
Römer, H. (**1993**). "Environmental and biological constraints for the evolution of long-range signalling and hearing in acoustic insects," Philos. Trans. R. Soc. London, Ser. B **340**, 179–185.
Schul, J. (**1998**). "Song recognition by temporal cues in a group of closely bushcricket species (genus *Tettigonia*)," J. Comp. Physiol. [A] **183**, 401–410.
Stoddard, P. K. (**1998**). "Application of filters in bioacoustics," in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J.Owren, and C. S. Evans (Springer, Berlin), pp. 105–127.

Sueur, J. (**2002**). "Cicada acoustic communication: Potential sound partitioning in a multispecies community from Mexico (Hemiptera: Cicadomorpha: Cicadidae)," Biol. J. Linn. Soc. **75**, 379–394.
Sueur, J., Aubin, T., and Simonis, C. (**2008a**). "Seewave: A free modular tool for sound analysis and synthesis," Bioacoustics **18**, 213–226.
Sueur, J., Windmill, J. F. C., and Robert, D. (**2006**). "Tuning the drum: The mechanical basis for frequency discrimination in a Mediterranean cicada," J. Exp. Biol. **209**, 4115–4128.
Sueur, J., Windmill, J. F. C., and Robert, D. (**2008b**). "Sexual dimorphism in auditory mechanics: Tympanal vibrations of *Cicada orni*," J. Exp. Biol. **211**, 2379–2387.
Windmill, J. F. C., Bockenhauer, S., and Robert, D. (**2008**). "Time-resolved tympanal mechanics of the locust," J. R. Soc., Interface **5**, 1435–1443.
Windmill, J. F. C., Jackson, J. C., Tuck, E. J., and Robert, D. (**2006**). "Keeping up with bats: Dynamic auditory in a moth," Curr. Biol. **16**, 2418–2423.
Windmill, J. F. C., Sueur, J., and Robert, D. (**2009**). "The next step in cicada audition: Measuring pico-mechanics in the cicada's ear," J. Exp. Biol. **212**, 4079–4083.

Erratum, Part II: Coupled vibration analysis of the thin-walled cylindrical piezoelectric ceramic transducers [J. Acoust. Soc. Am. 125, 803–818 (2009)]

Boris Aronov

Acoustics Research Laboratory, Advanced Technology and Manufacturing Center, Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723

(Received 15 December 2009; accepted 16 December 2009)

[DOI: 10.1121/1.3290998]

PACS number(s): 43.38.Ar, 43.10.Vx [AJZ]

This is an additional list of errors for the subject manuscript, and it supplements the list previously published [B. Aronov, "Erratum: Coupled vibration analysis of the thin-walled cylindrical piezoelectric ceramic transducers," J. Acoust. Soc. Am. **125**, 3467(E) (2009)].

Mistakes are found in Eqs. (17), (18), and (22). The corrected expressions should read as follows:

$$K_{03} = -\frac{4thY}{3a}. \tag{17}$$

$$\frac{d}{dt}\left(\frac{\partial W_{\text{kin}}}{\partial \dot{\xi}_i}\right) + \frac{\partial W_{\text{pot}}}{\partial \xi_i} = 0. \tag{18}$$

$$Z_{ii}U + \sum_{l \neq i} z_{li}U_l = 0. \tag{22}$$

 0001-4966/2010/127(3)/1689/1/$25.00

Erratum: "What makes a melody: The perceptual singularity of pitch sequences" [J. Acoust. Soc. Am. 126, 3179–3187 (2009)]

Marion Cousineau[a)]
Laboratoire Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes, and Ecole Normale Supérieure (Département d'Etudes Cognitivies), 29 rue d'Ulm, F-75230 Paris Cedex 05, France

Laurent Demany
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), Université de Bordeaux, BP 63, 146 rue Leo Saignat, F-33076 Bordeaux, France

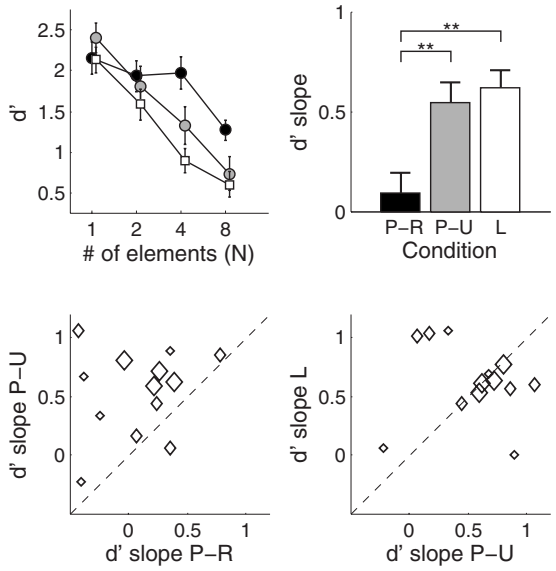
Daniel Pressnitzer
Laboratoire Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes, and Ecole Normale Supérieure (Département d'Etudes Cognitivies), 29 rue d'Ulm, F-75230 Paris Cedex 05, France

(Received 31 December 2009; accepted 5 January 2010)

[DOI: 10.1121/1.3298934]

PACS number(s): 43.66.Mk, 43.66.Hg, 43.10.Vx [BCM]

In the original paper, the bottom panels of Fig. 2 contain mistakes. The correct figure is presented here. The legend is unchanged.



[a)]Author to whom correspondence should be addressed. Electronic mail: marion.cousineau@ens.fr

 0001-4966/2010/127(3)/1690/1/$25.00

Erratum: Acoustic and perceptual similarity of Japanese and American English vowels [J. Acoust. Soc. Am. 124, 576–588 (2008)]

Kanae Nishi and Winifred Strange
Ph.D. Program in Speech and Hearing Sciences, Graduate School and University Center, City University of New York, New York, New York 10016

Reiko Akahane-Yamada and Rieko Kubo
Cognitive Information Science Laboratory, Advanced Telecommunications Research Institute International and ATR Learning Technology Corporation, Kyoto 619-0288, Japan

Sonja A. Trent-Brown
Department of Psychology, University of South Florida, Tampa, Florida 33620

(Received 17 November 2009; accepted 15 December 2009)

[DOI: 10.1121/1.3290993]

PACS number(s): 43.71.Hw, 43.70.Kv, 43.71.Es, 43.70.Fq, 43.10.Vx [MSS]

The caption for Table III on page 582 is incorrect. The correct caption is "TABLE III. Acoustic similarity (F1/F2 in Bark values) of Japanese and American English (AE) vowels: hVb(a) syllables produced in sentence."

 0001-4966/2010/127(3)/1691/1/$25.00

ACOUSTICAL NEWS

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

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

President's report on the 158th meeting of the Acoustical Society of America held in San Antonio, Texas

The 158th meeting of the Acoustical Society of America was held 26-30 October 2009 at the Hyatt Regency San Antonio Hotel in San Antonio, Texas. This was the first time that the Society has met in this city.

The meeting drew a total of 880 registrants, including 148 nonmembers, 234 students and 91 registrants from outside North America. There were 16 registrants from Japan; 13 from France; 8 from Korea; 7 from the UK; 6 from Russia; 5 each from Germany and Taiwan; 4 each from Chile and China; 3 each from Brazil, Norway, Singapore, and Sweden; 2 each from Italy and Portugal; and 1 each from Australia, Austria, Denmark, Finland, India, the Netherlands, and Thailand. North American countries, Canada, Mexico, and the United States, accounted for 22, 3, and 764 respectively.

A total of 660 papers, organized into 82 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included 10 meetings dealing with standards. The evening tutorial lecture series was continued by Kathleen P. King of Fordham University who presented "Podcasting Demystified: From Concept to Production for Every Budget."

A short course titled "Acoustic Imaging Applications: Seismic, Underwater, Biomedical, and Speech" was held with a group of 26 students. The instructors were Max Deffenbaugh, research geophysicist with ExxonMobil Research & Engineering Company, Jules Jaffe, Research Oceanographer at the Scripps Institution of Oceanography, Christy Holland, Associate Professor in the Department of Biomedical Engineering at the University of Cincinnati, and Maureen Stone, Professor at the University of Maryland Dental School.

A technical tour was organized to visit ETS-Lindgren Acoustic Research Laboratory and Factory. The tour included a technical discussion on acoustic test facilities followed by a factory tour. Participants toured several state-of-the-art chambers for acoustic test services, including a hemianechoic chamber and two reverberation chambers, impedance tubes and supporting acoustic test equipment and software.

The Society's thirteen Technical Committees held open meetings during the San Antonio meeting where they made plans for special sessions at upcoming ASA meetings, discussed topics of interest to the attendees and held informal socials after the end of the official business. These are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA Website http://asa.aip.org/committees.html, and in the Acoustical News section of JASA in the September and November issues.

The ASA Student Council hosted a Student Reception with over 100 people in attendance. This reception, which was sponsored by ASA and supported by the National Council of Acoustical Consultants, enabled students to meet with established members of the Acoustical Society of America. Several of the Technical Committees awarded Best Student Paper Awards or Young Presenter Awards to students and young professionals who presented papers at the meeting. The list of award recipients, as well as other information for students, can be found online at the ASA Student Zone website http://www.acosoc.org/student/.

Social events included the two social hours held on Tuesday and Thursday, an "icebreaker" and a reception for students, a Jam Session, the Fellows Luncheon and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions. The Women in Acoustics Luncheon was held on Wednesday afternoon with attendance of about 75. The Fellows Luncheon was attended by 100 people who heard a talk by Professor Todd Ditmire of the University of Texas at Austin Physics Department about very high energy lasers and some potential applications.

The plenary session included announcements, acknowledgment of the members and other volunteers who organized the meeting and the presentation of awards and certificates to newly-elected Fellows.

ASA President Whitlow Au presided over the Plenary Session and Awards Ceremony (see Fig. 1).

In September 2009, Whitlow Au and Charles Schmid traveled to Beijing to attend the 10th Western Pacific Acoustics Conference (WESPAC) and to meet with a delegation of Chinese acousticians to continue discussions to organize a future joint meeting of the ASA with the Acoustical Society of China. The Chinese delegation included Jing Tian (president of the Acoustical Society of China), Tom Ho (president of the Hong Kong Institute of Acoustics), Fengha Li (Institute of Acoustics China Academy of Sciences), and Renhe Zhang (Chinese representative to WESPAC). At that meeting, it was suggested that since WESPAC is scheduled to meet in Hong Kong in 2012, that ASA should consider making that meeting a joint ASA-WESPAC (which would include the Acoustical Society of China). Tom Ho and Maurice Kwok-leung attended the San Antonio meeting and made a presentation to the Executive and Technical Councils proposing a joint meeting. The Technical Council referred the proposal for discussion at each of the open technical committee meetings and favorable responses were received. Based on the Technical Council report, the Executive Council voted to move ahead with the joint meeting in Hong Kong, tentatively set for May of 2012.

Mardi Hastings, Chair of the Spring 2010 meeting, addressed the audience and invited and encouraged them to attend that meeting to be held in Baltimore, Maryland, in April (see Fig. 2).

Two ASA Science Writing Awards were presented. The 2008 Science Writing Award in Acoustics for Journalists was presented to Rachel Ehrenberg for her article "Stranded: A Whale of a Mystery" in *Science News*, July 2008 (see Fig. 3). The 2008 Science Writing Award for Professionals in Acoustics was presented to Ingo Titze for his article "The Human Instrument" in *Scientific American*, January 2008 (see Fig. 4).

FIG. 1. (Color online) Whitlow W.L. Au, ASA President.

 0001-4966/2010/127(3)/1692/7/$25.00

FIG. 2. (Color online) Mardi Hastings, Chair of the Spring 2010 meeting.

The ASA President introduced Kenneth Suslick, recipient of the 2009 Student Council Mentoring Award (see Fig. 5). He then introduced James F. Lynch, recipient of the Munk Award. The Munk Award and Medal are granted jointly by The Oceanography Society, the Office of Naval Research, and the Office of the Oceanographer of the Navy. There have been eight recipients since the award was established in 1993. This was the first time that the Munk Award was presented at an ASA meeting. James Lynch presented the Munk Award Lecture titled "Acoustical oceanography and shallow water acoustics" earlier in the meeting (see Fig. 6).

FIG. 4. (Color online) ASA President Whitlow Au (l) presents Science Writing Award for Professionals in Acoustics to Ingo R. Titze (r).

The 2009 Rossing Prize in Acoustics Education was presented to James F. Sanders, who is retired from the Naval Postgraduate School. The award includes a silver medal and a prize of $4,000. Dr. Sanders presented the Acoustics Education Prize Lecture titled "Fundamental acoustics education and applications" earlier in the meeting (see Fig. 7).

FIG. 3. (Color online) ASA President Whitlow Au (2) presents Science Writing Award in Acoustics for Journalists to Rachel Ehrenberg (r).

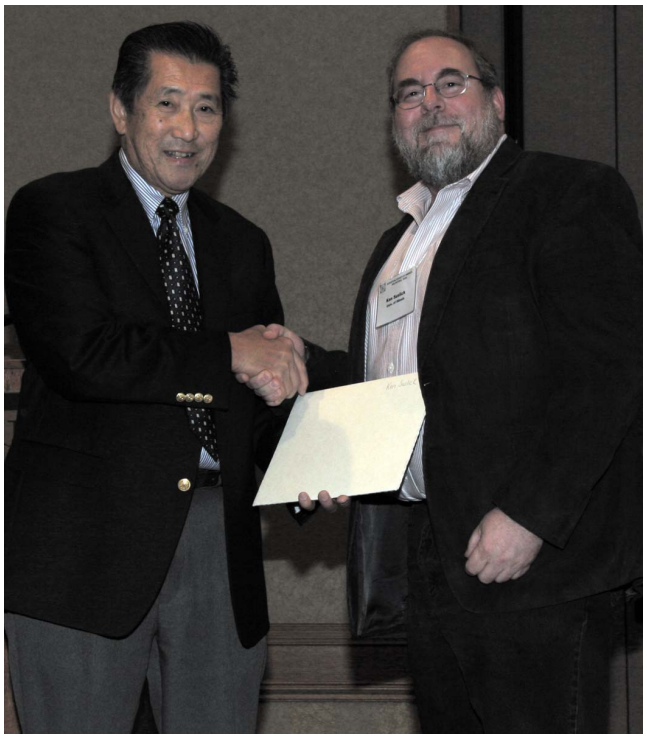


FIG. 5. (Color online) ASA President Whitlow Au (l) presents Student Council Mentoring Award to Kenneth Suslick (r).

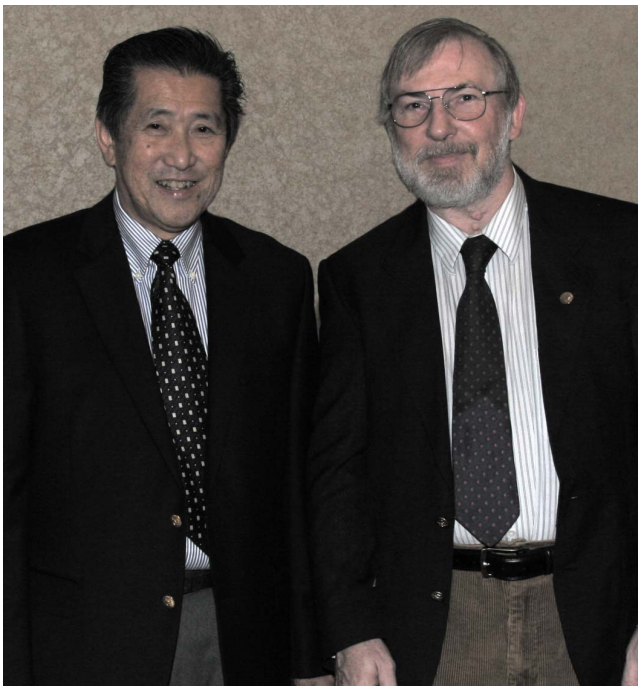


FIG. 6. (Color online) ASA President Whitlow Au (l) congratulates James F. Lynch, recipient of the 2009 Munk Award (r).

FIG. 8. (Color online) ASA President Whitlow Au (l) presents Silver Medal in Acoustical Oceanography to Robert C. Spindel (r).

The Silver Medal in Acoustical Oceanography was presented to Robert C. Spindel of the University of Washington "for implementation of ocean acoustic tomography and basin scale acoustic thermometry" (see Fig. 8). The Silver Medal in Noise was presented to Michael R. Stinson, National Research Council of Canada, "for contributions to outdoor sound propagation, acoustical materials, and ear canal acoustics" (see Fig. 9).

Election of twenty members to Fellow grade was announced and fellowship certificates were presented. New fellows are: David L. Adams, Michael A. Averkiou, Constantin C. Coussios, Diane Dalecki, Mostafa Fatemi, Kullervo H. Hynynen, Michael F. Insana, Boris G. Katsnelson, Robert G. Leisure, Plilipos C. Loizou, Patrick J. Loughlin, David A. Mann, David K. Mellinger, Ralph T. Muehleisen, Susan N. Nittrouer, Dennis A. Paoletti, Kevin B. Smith, Andrew J. Szeri, Shin-ichiro Umemura, and Oleg A. Sapozhnikov (see Fig. 10).

ASA President Whitlow Au introduced Clark Penrod, Chair of the San Antonio meeting (see Fig. 11) who expressed the Society's thanks to mem-

FIG. 7. (Color online) ASA President Whitlow Au (l) presents Rossing Prize in Acoustics Education to James V. Sanders (r).

FIG. 9. (Color online) ASA President Whitlow Au (l) presents Silver Medal in Noise to Michael R. Stinson (r).

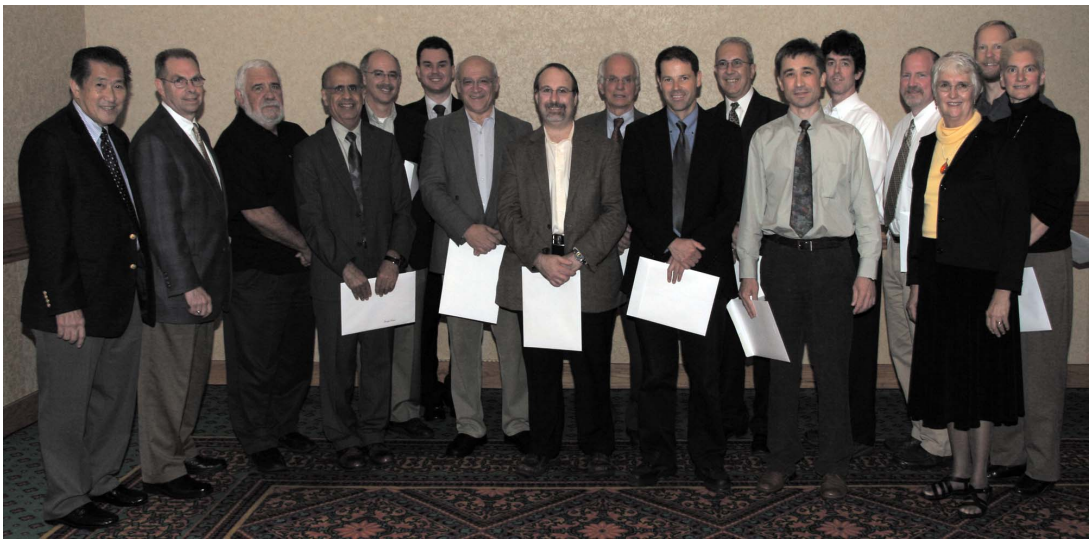


FIG. 10. (Color online) New Fellows of the Acoustical Society of America with ASA President Whitlow Au and ASA Vice President Diane Kewley-Port (l to r): Whitlow W.L. Au, Dennis A. Paoletti, David L. Adams, Mostafa Fatemi, Michael F. Insana, Constantin C. Coussios, Boris Katsnelson, Ralph R. Muehleisen, Robert G. Leisure, Andrew J. Szeri, Philip Loizou, Oleg A. Sapozhnikov, Patrick J. Loughlin, Kevin B. Smith, Diane Kewley-Port, David K. Mellinger, Diane Dalecki.

FIG. 11. (Color online) Clark Penrod, Chair of the San Antonio meeting.

bers of the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning, including Clark S. Penrod, General Chair; Nicholas P. Chotiros, Technical Program Chair; Tim Hawkins, Deputy to the General Chair; Jim Stockton, Assistant to the Deputy; Leonard Hebert, Audio/Visual; Debbie Smith Matthys, Food and Beverage; Patti Neary, Registration; Dixie Terry, Grace Rau, General Assistance; Judy Stockton, Accompanying Persons Program; Mark Hamilton, Preston Wilson, Mark Wochner, Mike Pestorius, Advisors to the Local Committee (see Fig. 12). He also expressed thanks to the members of the Technical Program Organizing Committee: Nicholas P. Chotiros, Chair; Juan I. Arvelo, Acoustical Oceanography; Rama Ratnam, Animal Bioacoustics; Douglas F. Winker, David S. Woolworth, Architectural Acoustics; Mark Wochner, Biomedical Ultrasound/Bioresponse to Vibration; Preston S. Wilson, Education in Acoustics and Physical Acoustics; Daniel Costley, Engineering Acoustics; James M. Cottingham, Musical Acoustics; Pamela J. Harght, Noise; Craig Champlin, Psychological and Physiological Acoustics; Natalia Sodorovskaia, Signal Processing in Acoustics; Rajka Smiljanic, Augustine Agwuele, Speech Communication; Kai Ming Li, Structural Acoustics and Vibration; Marcia J. Isakson, Underwater Acoustics (see Fig. 13).

The full technical program and award encomiums can be found in the printed meeting program or online for readers who wish to obtain further information about the San Antonio meeting (visit scitation.aip.org/jasa and select Volume 12, Issue 4, from the list of available volumes).

FIG. 12. (Color online) Meeting Organizing Committee (l to r): Leonard Hebert, Tim Hawkins, Judy Stockton, Jim Stockton, Mike Pestorius, Debbie Smith-Mathys, Nicholas Chotiros, Clark Penrod.

FIG. 13. (Color online) Technical Program Organizing meeting (l to r): Nicholas Chotiros, Natalie Sodorovskia, Kai Ming Li, Augustine Agwuele, Dan Costley, James Stockton, Mark Wochner, Marcia Isakson, Doug Winker, Tim Hawkins, Pamela Hargt, David Woolworth, Rajka Smiljanic, Preston Wilson, Juan Arvelo, Craig Champlain, James Cottingham, Judy Cottingham (not pictured Rama Ratnam). Photo by C. Schmid.

We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

WHITLOW W.L. AU
President 2009–2010

Regional Chapter News

Madras Regional Chapter

The Madras Regional Chapter of the Acoustical Society of America (MIRC-ASA) held three meetings in 2009, on 27 June, 8 August, and 26–27 November. The meeting in November was held jointly with the Acoustical Society of India (ASI). Silver and Gold Medal awards were presented by the International Research Institute for the Deaf (IRID), a unit of the Acoustical Foundation (AFECT).

At the 27 June meeting, Dr. P. Iyamperumal, Executive Director of Tamil Nadu Science and Technology Center (TNSTC), State Govt. of Tamil Nadu, Chennai, delivered an invited lecture titled "Architect of Burlap Planetarium" which reviewed the basics of acoustics. A large number of students and teachers participated in the meeting and 33 student projects were received.

The MIRC-ASA/IRID Acoustics Fair ceremony was held at TNSTC on 8 August. The following students in the junior group received awards from MIRC-ASA: Ms. B. Divya, Ms. S. Dixitha, Sri T. M. Krishnakumar (First place), Sri G.R. Ratheesh, Sri V. Srinivas, Sri K.E. Srivatsav, Ms. N. Venkatalakshmi., Sri V. Vaasanthosh, and Sri V.L. Venkatavaradhan (see

FIG. A1. (Color online) Student award presentation ceremony: (l to r) H. S. Paul (Chapter President), S. Dixita, N. Venkatalakshmi. V. Srinivas, G. R. Ratheesh, T. M. Krishnakumar, V. L. Venkatavadhan, K. E. Srivatsav, B. Divya, V. Vaasanthosh, R. Dhilsha (Chapter Treasurer).

FIG. A2. (Color online) First place student award presentation ceremony: (l to r) P. Iyamperumal (E.D., TNSTC), T. M. Krishnakumar, H. S. Paul (Chapter President).

FIG. A3. (Color online) H. S. Paul (l) presents Gold Medal to Dr. V. Bhujanga Rao (r).

FIG. A4. (Color online) Student award presentation ceremony: (l to r) V. Bhujanga Rao (ASI President), A. Ramachandraiah (ASI Vice President), S. Fatima (award recipient), H. S. Paul (Chapter President), S. K. Ray (Symposium Chair), and N.S. Naidu (ASI General Secretary).

FIG. A5. (Color online) (l to r) V. Bhujanga Rao (ASI President), A. Ramachandraiah (ASI Vice President), G. Rajkumar (award winner), H. S. Paul (Chapter President), S. K. Ray (Symposium Chair), and N.S. Naidu (ASI General Secretary).

FIG. A6. (Color online) (l to r) V. Bhujanga Rao (ASI President), A. Ramachandraiah (ASI Vice President), G. S. H. Gautham (award winner), H. S. Paul (Chapter President), S. K. Ray (Symposium Chair), and N.S. Naidu (ASI General Secretary).

Figs. A1 and A2). Awards were presented by H.S. Paul (President and Chapter Representative of MIRC-ASA and President of AFECT & IRID) and R. Dhilsha (Treasurer of MIRC-ASA). The MIRC-ASA Chapter expresses its gratitude to Dr. P. Iyamperumal, E.D. of the Tamil Nadu Science and Technology Center who paid for all expenses of the fair and provided all facilities for the first meeting.

The joint meeting between Acoustical Society of India (ASI) and MIRC-ASA was held 26-28 November 2009 at the Research Center Imarat (RCI), DRDO, Hyderabad, A.P. Sri K. Jagadisan (Convener of the Symposium) and Sri S. K. Ray (Chairman of the Symposium), Director of RCI, DRDO, Hyderabad, A.P. made arrangements for the MIRC-ASA chapter meeting to be incorporated into the main ASI meeting to strengthen the mutual bond between ASI and MIRC-ASA and to encourage the spread of acoustical science throughout India.

The MIRC-ASA and IRID (AFECT) award ceremony was conducted during the ASI opening session. Prof. A. Ramachandraiah (Vice President of MIRC-ASA and Trustee of AFECT) read the citation for Gold Medal recipient Dr. V. Bhujanga Rao, Director and Outstanding Scientist of Naval Science and Technological Laboratory (NSTL), Visakhapatnam, A.P. and President of ASI (see Fig. A3). The Gold Medal award is a joint venture of MIRC-ASA & IRID. After the opening session, Dr. Rao delivered the Mira Paul Memorial Distinguished Lecture on "Design of Indigenous Affordable Cochlear Implant System."

The following students were awarded MIRC-ASA best paper awards at the November meeting: Ms. S. Fatima, Sri G. Rajkumar, Sri G.S.H. Gowtha (see Figs. A4–A6). The following students also received awards in absentia Sri M. M. Armstrong Arasu, Ms. Lakshimi Venkatesh, and Sri E. Rajasekar.

H. S. Paul

Calendar of Meetings and Congresses

Compiled by the Information Service of the International Commission for Acoustics

2010

8–10 March	Chofu, Japan. Spring Meeting of the Acoustical Society of Japan.
15–18 March	Berlin, Germany. Meeting of the German Association for Acoustics DAGA 2010. http://www.daga-tagung.de/2010
15–19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and SignalProcessing. http://icassp2010.org
7–9 April	Cambridge, UK. David Weston Sonar Performance Assessment Symposium
19–23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society ofAmerica and Noise-Con 2010. http://asa.aip.org/meetings.html
27–30 April	Ghent, Belgium. Institute of Acoustics/Belgian Acoustical Association JointMeeting. http://www.ioa.org.uk/viewupcoming.asp
6–7 May	Paris, France. 2nd International Symposium on Ambisonics and SphericalAcoustics. http://ambisonics10.ircam.fr
10–12 May	Bergen, Norway. Baltic-Nordic Acoustics Meeting 2010. E-mail: sanordby@norsonic.com
9–11 June	Aalborg, Denmark. 14th Conference on Low Frequency Noise and Vibration. http://lowfrequency2010.org
13–16 June	Lisbon, Portugal. INTERNOISE2010. http://www.internoise2010.org
5–9 July	Istanbul, Turkey. 10th European Conference on Underwater Acoustics. http://ecua-2010-istanbul.org
18–22 July	Cairo, Egypt. 17thh International Congress on Sound and Vibration (ICSV17). http://www.icsv17.org
23–27 August	Sydney, Australia. International Congress on Acoustics 2010.http://www.ica2010sydney.org
23–27 August	Seattle, USA. 11tth International Conference on Music Perception and Cognition.
29–31 August	Melbourne, Australia. International Symposium on Room Acoustics (ISRA2010). http://www.isra2010.org
14–18 September	Kyoto, Japan. 5th Animal Sonar Symposium. http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–18 September	Ljubljania, Slovenia. Alps-Adria-Acoustics Association Meeting jointwith EAA. E-mail: mirko.cudina@fs.uni-lj.si
20–22 September	Leuven, Belgium, International Conference on Noise and VibrationEngineering (ISMA2010). http://www.isma-isaac.be
26–30 September	Makuhari, Japan. Interspeech 2010—ICSLP. http://www.interspeech2010.org
11–14 October	San Diego, Cal. USA. IEEE 2010 Ultrasonics Symposium. E-mail: b.potter@vectron.com
14–16 October	Niagara-on-the Lake, Ont., Canada. Acoustics Week in Canada. http://caa-aca.ca/E/index.html
11–14 October	San Diego, California, USA. IEEE 2010 Ultrasonics Symposium. E-mail: bpotter@vectron.com
13–15 October	Leon, Spain. 41st Spanish Congress of Acoustics and 6th Iberian Acoustics Congress. http:// www.sea-acustica.es
18–22 October	Nagahama, Japan. 10th International Workshop on Railway Noise (IWRN10). http://www.rtri.or.jp/IWRN10/first.announcement.html
15–19 November	Cancún, Mexico. 2nd Pan-American Meeting on Acoustics. http://asa.aip.org/meetings.html

2011

22–27 May	Prague, Czech Republic. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2011). http://www.icassp2011.com
23–27 May	Seattle, WA, USA. 160th meeting of the Acoustical Society of America. http://asa.aip.org/meetings.html
27 June–1 July	Aalborg, Denmark. Forum Acusticum 2011. http://www.fa2011.org
24–28 July	Tokyo, Japan. 19th International Symposium on Nonlinear Acoustics (ISNA 19). http://www.isna19.com
27–31 August	Florence, Italy. Interspeech 2011. http://www.interspeech2011.org
4–7 September	Osaka, Japan. Internoise 2011. http://www.internoise2011.com
5–8 September	Gdansk, Poland. International Congress on Ultrasonics. http://icu2011.ug.edu.pl/index.html

2012

20–25 March	Kyoto, Japan. IEEE International Conference on Acoustics, Speech, and Signal Processing. http://www.icssp2012.com

2013

26–31 March	Vancouver, Canada. 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). http://www.icassp2013.com
2–7 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013). http://www.ica2013montreal.org

ACOUSTICAL STANDARDS NEWS

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

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

Accredited Standards Committees S1, Acoustics; S2, Mechanical Vibration and Shock; S3, Bioacoustics; S3/SC 1, Animal Bioacoustics; and S12, Noise, along with the U.S. Technical Advisory Groups to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its five Subcommittees, and IEC/TC 29, ASACOS and the Standards Plenary Group will meet in conjunction with the Joint 159th ASA Meeting and Noise-Con 2010 to be held in Baltimore, Maryland, 19-23 April 2010. Specific meeting dates and times follow:

• Tuesday, 20 April 2010

– **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.

• Thursday, 22 April 2010

– **8:00 a.m.—9:00 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**.

– **9:15 a.m.—10:30 a.m.**—Meeting of Accredited Standards Committee **S12** on Noise

– **11:00 a.m.—12:00 p.m.**—Meeting of Accredited Standards Committee **S2** on Mechanical Vibration and Shock

– **1:45 p.m.—2:45 p.m.**—Meeting of Accredited Standards Committee **S1** on Acoustics

– **3:00 p.m.—4:15 p.m.**—Meeting of Accredited Standards Committee **S3** on Bioacoustics

– **4:30 p.m.—5:30 p.m.**—Meeting of Accredited Standards Committee **S3/SC 1** on Animal Bioacoustics

Standards Meetings Calendar—International

• 20–24 September 2010

– **ISO/TC 108/SC 4 and ISO/TC 108/SC 5**—London, UK

• 27 September–1 October 2010

– **ISO/TC 108** and its subcommittees **ISO/TC 108/SC 2, SC 3**, and **SC 6**, London, UK

For additional information on these meetings, please contact the Standards Secretariat.

Thanks to our Volunteers!

The ASA Standards Secretariat would like the thank the following people for volunteering their time and expertise to coordinate comments and formulate recommendations for the U.S. vote on numerous ISO and IEC documents throughout the second half of the past year. Their efforts are greatly appreciated and extremely valuable. The following is a list of volunteers from July 2009 through the end of the year.

K. Bever
R. Dong
B.E. Douglas
N.H.W. Eklund
W.C. Foiles
R.L. Grason
L. Hawkins
B. Lang
L.A. Wilber
S. Lind
W. Madigosky
A.H. Marsh
E.H. Maslen
L. Pater
D.A. Preves
P.D. Schomer
D.J. Vendittis

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 S12) (Acoustical Society of America)

Revisions

BSR/ASA S12.42-201x, Methods for the Measurement of Insertion Loss of Hearing Protection Devices in Continuous or Impulsive Noise Using

Microphone-in-Real-Ear or Acoustic Test Fixture Procedures (revision and redesignation of ANSI S12.42-1995 (R2004))

Provides two methods for measuring the insertion loss of any hearing protection device (HPD) that encloses the ears, caps the ears, or occludes the earcanals. This standard contains information on instrumentation, calibration, electroacoustic requirements, subject selection and training, procedures for locating ear-mounted microphones and HPDs to measure sound pressure levels at the ear, specifications describing suitable ATFs, and methods for reporting the calculated insertion-loss values. (December 28, 2009)

ASME (American Society of Mechanical Engineers)

Revisions

BSR/ASME BPVC Section V-201x, Nondestructive Examination (11/05/09 Meeting) (revision of ANSI/ASME BPVC 2007 Edition)

Contains requirements and methods for nondestructive examination (NDE), which are referenced and required by other Sections of the Code. These NDE methods are intended to detect surface and internal imperfections in materials, welds, fabricated parts and components. The following NDE methods are addressed: radiography; ultrasonics; liquid penetrant; magnetic particle; eddy current; visual; leak testing; and acoustic emission. (February 1, 2010)

ATIS (Alliance for Telecommunications Industry Solutions)

Reaffirmations

BSR ATIS 0300221-1995 (R201x), OAM&P—In-Service, Nonintrusive Measurement Device (NMD) Voice Service Measurements (reaffirmation of ANSI ATIS 0300221-1995 (R2004))

Provides specifications for in-service nonintrusive measurement devices (INMD) used to measure various parameters of importance to voice service transmission maintenance of telecommunications networks. These measurement devices are used primarily for the measurement of voicegrade analog parameters such as speech level, noise level, echo path loss and echo path delay. This standard specifies interface, measurement range and accuracy requirements for measuring voicegrade transmission parameters as well as descriptions of optional functions associated with these parameters. (January 18, 2010)

TIA (Telecommunications Industry Association)

Revisions

BSR/TIA 470.120-C-201x, Telecommunications—Telephone Terminal Equipment—Transmission Requirements for Analog Speakerphones (revision and redesignation of ANSI/TIA 470-B-1997)

Provides speakerphone acoustic performance requirements for Customer Premises Equipment (CPE) intended for analog connection to the Public Switched Telephone Network (PSTN). These requirements should ensure compatibility and satisfactory performance to the user in a high percentage of installations. Test measurement methods reference procedures in IEEE Std 1329, where applicable. (January 11, 2010)

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.

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

New Standards

ANSI/AHRI Standard 270-2009, Sound Performance Rating of Outdoor Unitary Equipment (new standard)

ANSI/AHRI Standard 300-2009, Sound Rating and Sound Transmission Loss of Packaged Terminal Equipment (new standard)

ANSI/AHRI Standard 350-2009, Sound Performance Rating of Non-Ducted Indoor Air-Conditioning Equipment (new standard)

ASA (ASC S2) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S2.31-1979 (R2009), Methods for the Experimental Determination of Mechanical Mobility—Part 1: Basic Definitions and Transducers (reaffirmation and redesignation of ANSI S2.31-1979 (R2004))

ANSI/ASA S2.32-1982 (R2009), Methods for the Experimental Determination of Mechanical Mobility—Part 2: Measurements Using Single-Point Translational Excitation (reaffirmation and redesignation of ANSI S2.32-1982 (R2004))

ASA (ASC S12) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S12.53/Part 2-1999 (R2009)/ISO 3743-2-1994 (R2009), Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberant fields—Part 2: Methods for special reverberation test rooms (reaffirmation and redesignation of ANSI S12.53/Part 2-1999/ISO 3743-2-1994 (R2004))

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.

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.41-201x, Audible Emergency Notification and Evacuation Signal (revision and redesignation of ANSI ASA S3.41-1990 (R2008))

Applies to an audible emergency notification and evacuation signal. This standard specifies three parameters of the signals, i.e., temporal pattern, sound pressure level, and intelligibility (in voice systems as required) at all places within the intended signal reception. The spectral content of the signal is not specified and should be selected to satisfy specific site requirements and/or regulations. It applies to the audible signal and not to the individual signaling system components. Project Need: To broaden this standard owing to recent consideration for addressing the unique needs of handicapped individuals and the increased concern and need for heightened personal safety. Stakeholders: DOD, telecomunications industry, educators, administrators, manufacturing facilities management.

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 3822-3/Amd1:2009, 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—Amendment 1

ISO 8253-2:2009, Acoustics—Audiometric test methods—Part 2: Sound field audiometry with pure-tone and narrow-band test signals

COMPRESSORS, PNEUMATIC TOOLS AND PNEUMATIC MACHINES (TC 118)

ISO 28927-1:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 1: Angle and vertical grinders

ISO 28927-2:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 2: Wrenches, nutrunners and screwdrivers

ISO 28927-3:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 3: Polishers and rotary, orbital and random orbital sanders

ISO 28927-5:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 5: Drills and impact drills

ISO 28927-6:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 6: Rammers

ISO 28927-7:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 7: Nibblers and shears

ISO 28927-8:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 8: Saws, polishing and filing machines with reciprocating action and small saws with oscillating or rotating action

ISO 28927-9:2009, Hand-held portable power tools—Test methods for evaluation of vibration emission—Part 9: Scaling hammers and needle scalers

ISO and IEC Draft International Standards

ISO

Acoustics (TC 43)

ISO/DIS 3382-3, Acoustics—Measurement of room acoustic parameters—Part 3: Open plan spaces (2/25/2010)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 7626-1, Vibration and shock—Experimental determination of mechanical mobility—Part 1: Basic definitions and transducers (2/1/2010)

ISO/DIS 13379-1, Condition monitoring and diagnostics of machines—Data interpretation and diagnostics techniques—Part 1: General guidelines (1/21/2010)

IEC Draft Standards

59/546/FDIS, IEC 60704-1 Ed 3.0: Household and similar electrical appliances—Test code for the determination of airborne noise—Part 1: General requirements, (01/29/2010)

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*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
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*

7,587,826

43.10.Pr HAND HELD VIBRATING KNIFE

John David Stutz, Suffolk, Virginia
15 September 2009 (Class 30/142); filed 17 January 2006

A knife as described in this patent is made to vibrate by a motor-driven imbalanced rotor attached near the handle. The patent makes no mention of user-adjusted controls or of counter-rotating elements for changing the direction and/or magnitude of the vibrations. It does mention that the rotating elements generate a gyroscopic effect that may be used to stabilize the blade during the cutting process, but provides no hint of how this unlikely benefit may be achieved in practice.—EEU

7,598,640

43.10.Pr VIBRATION EXCITER

Christian Heichel *et al.*, assignors to ABI Anlagentechnik-Baumaschinen-Industriebedarf Maschinenfabrik und Vertriebsgesellschaft mbH
6 October 2009 (Class 310/81); filed in Germany 7 March 2007

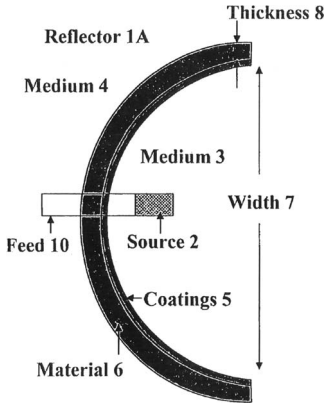
This vibration generator, intended for use in vibratory pile driving and the like, is based on the time-honored configuration consisting of imbalance masses on counter-rotating shafts. A rotary hydraulic oscillating motor, built into one of the shafts, is provided for changing the relative position of these masses. A disk brake can lock the arrangement to the housing in a desired orientation.—EEU

7,593,289

43.35.Zc REFLECTORS AND REFLECTOR LIGHT AND SOUND SOURCE SYSTEMS

Raymond B. Schaefer *et al.*, assignors to Phoenix Science & Technology, Incorporated
22 September 2009 (Class 367/138); filed 17 April 2006

According to this patent, a variety of industrial and military applications make use of concentrated light and/or sound—paint stripping, for example. Sound can be generated by a spark or other source that delivers substantial energy in the ultrasonic range. Several embodiments are described, but they all make use of a multi-layer reflector. The material and



thickness of the base layer 6 are chosen for efficient sound reflection, and coatings 5 are chosen for efficient light reflection.—GLA

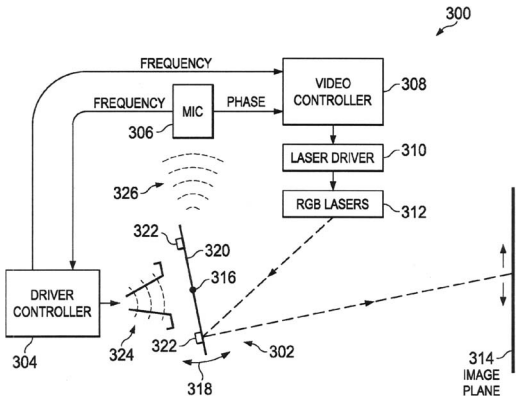
7,586,659

43.38.Ar AUDIO MEMS MIRROR FEEDBACK

Arthur Monroe Turner and Varkey George, assignors to Texas Instruments Incorporated
8 September 2009 (Class 359/199.1); filed 16 January 2009

This curious patent discloses the use of a microelectronic mechanical (MEMS) microphone to detect the motions of a MEMS mirror device of the type that is used in projection televisions. The only complete representation of the mechanism being patented is shown in the figure included here. The patent disclosure is not very detailed in its description, but it appears that the microphone is used to detect the velocity of air pulses coming from the edge of the moving mirror. The patent appears to be for a concept rather than a working device, as the authors make no mention of the fact that the majority of the signal from the microphone in this position would be at the second harmonic of the motion frequency, and they mention electromagnetic actuation of the mirrors as a possibility when there are no MEMS-scale solenoids available at this time, at least to this reviewer's knowledge. One would also

 0001-4966/2010/127(3)/1702/10/$25.00



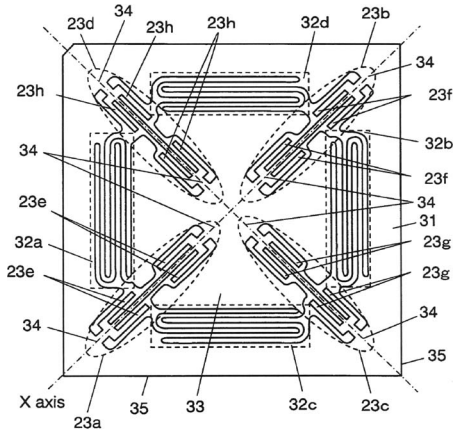
question the wisdom of using air as a coupling medium within a feedback loop, with the possibility of external interference and atmospheric effects affecting the response of the system.—JAH

7,587,941

43.38.Ar VIBRATION PIEZOELECTRIC ACCELERATION SENSOR

Jiro Terada *et al.*, assignors to Panasonic Corporation
15 September 2009 (Class 73/514.29); filed in Japan 2 March 2004

This patent discloses the use of a suspended plate as a mass for acceleration sensing. The plate is instrumented with piezoelectric strain gauges of a peculiar design situated at all four corners of the plate, as shown in the figure. These strain gauges have two active elements, one on either end of a silicon bar. They are used in transmitter/receiver pairs, each coupled to an



accompanying closed loop electromechanical resonator circuit. Apparently the inventors intend to use the transverse vibrations of the suspending bars as the sensing principle for they disclose the use of frequency to voltage converters in the detection electronics associated with the patented device. This implies that they rely on the tension in the bars as the frequency-determining factor, as opposed to the flexural stiffness of the bars. They apparently have the capability to fabricate very slender bars for this to work as illustrated.—JAH

7,598,655

43.38.Ar PIEZOELECTRIC TRANSFORMER DRIVING APPARATUS AND PIEZOELECTRIC TRANSFORMER DRIVING METHOD

Yasuhide Matsuo and Akira Mizutani, assignors to Tamura Corporation
6 October 2009 (Class 310/318); filed in Japan 6 October 2003

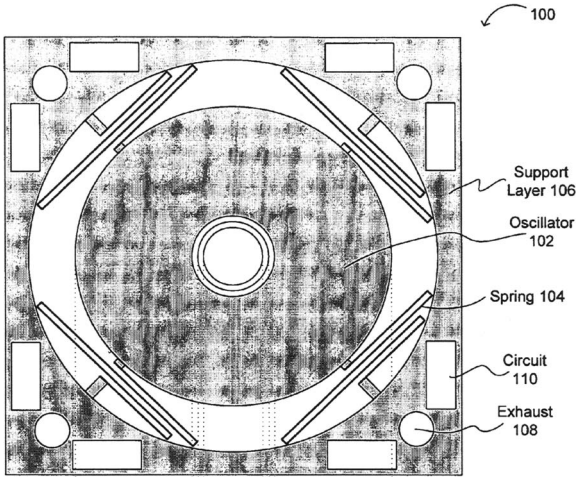
This patent discloses the use of Rosen-type piezoelectric transformers to drive high-voltage electronic devices differentially from low-voltage ac. The principle is fairly well established, but this patent provides a useful introduction to anyone interested in the details of the implementation.—JAH

7,570,775

43.38.Dv MICROELECTROMECHANICAL SPEAKER

Shinichi Araki, assignor to Sony Corporation
4 August 2009 (Class 381/191); filed 3 March 2005

What appears to be a very small electrodynamic loudspeaker of the moving magnet type is disclosed. Oscillator 102 (commonly called a diaphragm) is supported by springs 104. Exhaust ports 108 relieve back pressure or tune the back chamber. The coil and magnetic material are located in



the center of the diaphragm and are supported by a rigid layer below 102. How the coil and magnetic material are fabricated are briefly described. No device dimensions are noted, nor are the sensitivity or frequency response, for the device.—NAS

7,586,241

43.38.Fx ELECTROACOUSTIC TRANSDUCER

Hideaki Kuroda and Yoshihiro Sonoda, assignors to Murata Manufacturing Company, Limited
8 September 2009 (Class 310/348); filed in Japan 24 November 2005

The authors disclose the use of a flexible film substrate for mounting piezoelectric bimorph elements on in order to increase its radiating efficiency into air. The elements are disposed bilaterally in pairs or quadrilaterally around the film, which is supported by a rigid frame. The elements are rather large compared to the dimensions of the frame that supports the flexible film, which leads one to believe that the authors have not tried to optimize the size of and/or placement of the transducers. There is no discussion of the actual efficiency improvement over the bare transducers when radiating into air, but in reality one must trade off the improvement against the cost and complexity of fabricating and using such a hybrid structure.—JAH

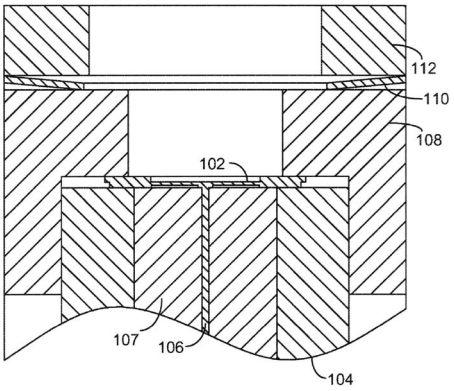
7,561,277

43.38.Gy MEMS FIBER OPTIC MICROPHONE

Ken K. Chin *et al.*, assignors to New Jersey Institute of Technology
14 July 2009 (Class 356/480); filed 18 May 2007

Micro-electric mechanical systems Fabry–Pérot diaphragm 102 is clamped to stainless steel ferrule 104 of single mode fiber 106, which is enclosed by zirconia ferrule 107, by washer 108, disk spring 110, and win-

dow cap 112. The patent states that the optic power output versus applied pressure of the device is an Airy function (how this is detected and decoded



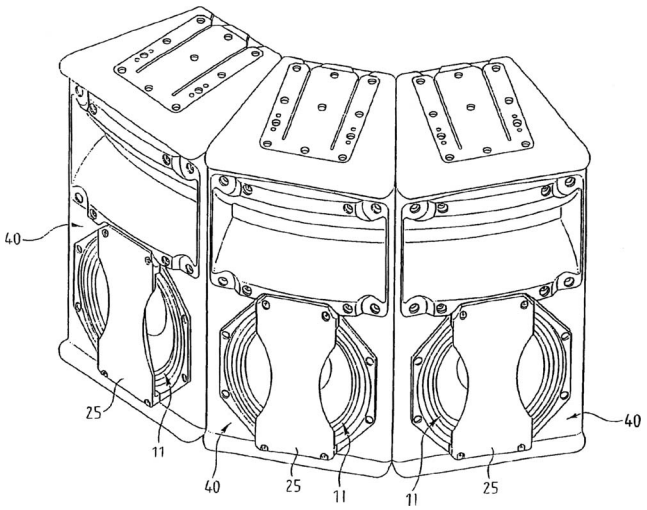
at the other end of the fiber is not described), and that multiple light sources are compatible with the device, including distributed feedback single mode lasers as well as (lower cost) LEDs. The description in the patent is concise and informative.—NAS

7,596,236

43.38.Hz LOUDSPEAKER WITH DIRECT EMISSION AND OPTIMISED RADIATION

Eric Vincenot, assignor to Nexo
29 September 2009 (Class 381/160); filed in France 23 July 2001

The illustration shows a typical curved line array, with cone loudspeakers 11 reproducing low and mid frequencies. If the speakers are, say, 30 cm apart then pattern control "explodes" above 550 Hz, whereas the high frequency horns operate efficiently only above 1 kHz or so. To get around



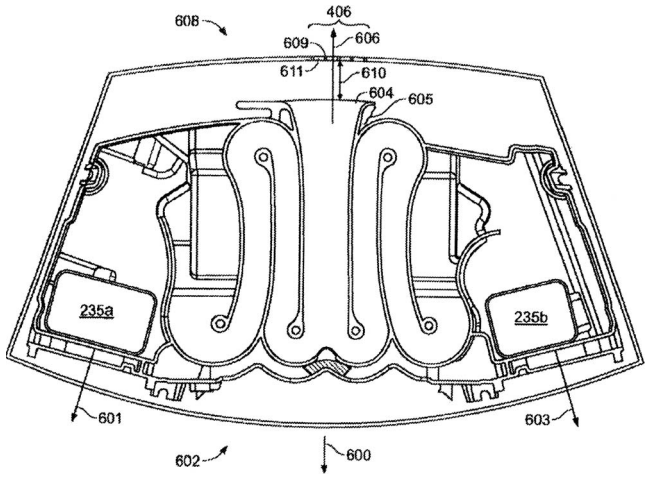
this problem some manufacturers use a pair of smaller woofers in each cabinet rather than a single larger speaker. The penalty is a reduction in cone area of about 50%. For many years, Nexo has employed a clever alternative: sound from each woofer exits through two horizontal slots, one above the other. In this improved version, a single "blocking member" 25 is attached to a conventional front-mount woofer.—GLA

7,584,820

43.38.Ja ACOUSTIC RADIATING

Robert Preston Parker and Dewey Potter, assignors to Bose Corporation
8 September 2009 (Class 181/155); filed 9 August 2004

The patent text includes some interesting ideas about loudspeakers mounted on tuned pipes, but few of these are actually included in the patent claims. The preferred embodiment consists of a pair of forward-facing loudspeakers 601 and 603 whose rear radiation passes through folded



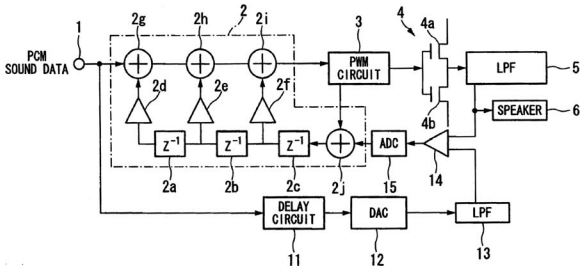
waveguides joined to a common exit 604. For this geometry to work reliably in a stereo system, channel separation should be progressively reduced below 200 Hz or so to provide monophonic bass drive, but this is not mentioned in the patent.—GLA

7,586,370

43.38.Lc CLASS D AMPLIFIER

Morito Morishima, assignor to Yamaha Corporation
8 September 2009 (Class 330/251); filed in Japan 18 March 2005

This is an interesting class D amplifier circuit. Also, the patent itself is interesting in that ten embodiments are described at some length yet the patent claims are few and short, with only a single independent claim. All



embodiments include two pulse width modulation circuits, complete with output lowpass filters 13 and 41. The two lowpassed analog signals are compared in summing amplifier 46, and the resulting error signal is used to provide corrective feedback.—GLA

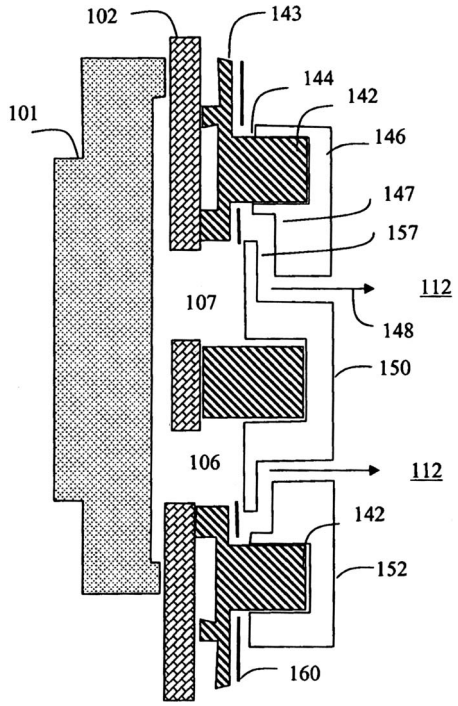
7,590,234

43.38.Si KEYPAD WITH FLANGED KEYS FOR HIGH AUDIO FRONT PORTING COMMUNICATION PRODUCTS

Donald W. Burnette *et al.*, assignors to Motorola, Incorporated
15 September 2009 (Class 379/433.07); filed 18 November 2005

Unlike Yamaha's U.S. Patent No. 7,586,370 which condenses ten distinct embodiments into four short claims, this patent manages to expand a one-sentence concept into 18 patent claims. In the effort to provide more and more functions in less and less space, some cellular phones use the keypad as a loudspeaker grill. However, current "zero-gap" keypads have no rigid separators between keys. As a result, any holes in the back membrane

can be seen by the user. Horrors! The remedy disclosed here is to use



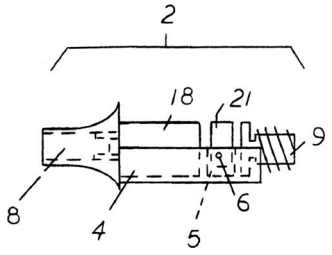
flanged keys, providing a concealed air path from speaker 101 to the outside world, as indicated by arrows 148.—GLA

7,590,258

43.38.Si IN-EAR EARPHONE

Mark Andrew Krywko, Bradenton, Florida
15 September 2009 (Class 381/380); filed 5 July 2007

Every component of this earphone is adjustable, removable, or replaceable. Ear tip 8, for example, can be selected for preferred treble response. The patent claims include two odd requirements: the receiver must



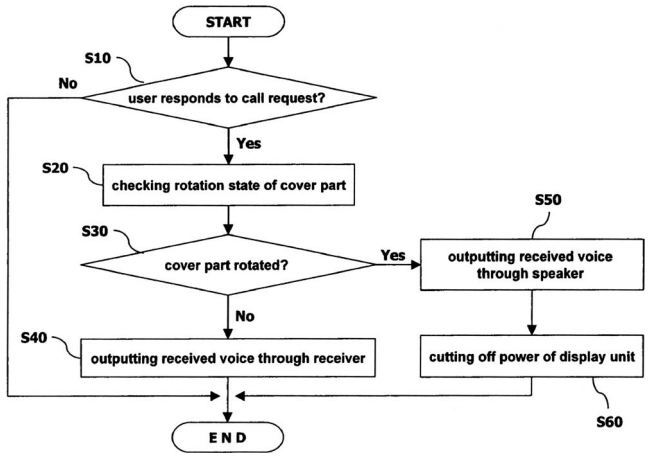
be a "single armature speaker" and the connecting cable must be coaxial.—GLA

7,590,431

43.38.Si SYSTEM AND METHOD FOR OUTPUTTING DUAL VOICE OF MOBILE TERMINAL

Hun-Soo Lee, assignor to LG Electronics Incorporated
15 September 2009 (Class 455/569.1); filed in Republic of Korea 27 November 2003

Cellular phones must be heard whether the case is open or closed, and this seemingly simple requirement is the basis for more than a half-dozen patents. In some cases, adjustable waveguides are used to control the output of a single loudspeaker; in others, the electrical signal is switched between two speakers. This patent belongs to the second category. As might be expected, it has become increasingly difficult to tiptoe around prior art, and a note of desperation has crept into the patent claims. Among other things, it is necessary that "…the swivel hinge rotates in a direction perpendicular to a front surface of the folder part," and "…at least one of the different output



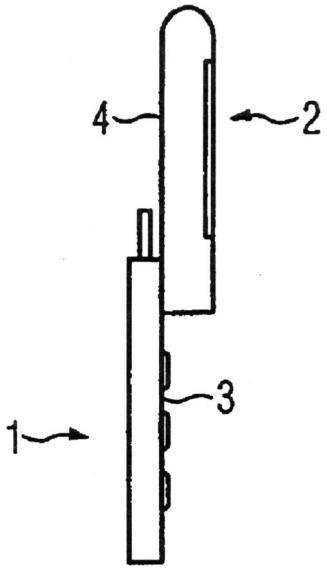
devices is controlled with reference to a sound pressure conversion table stored in memory."—GLA

7,590,436

43.38.Si SLIDER MOBILE TELEPHONE COMPRISING A BENDING WAVE LOUDSPEAKER

Michael Hülskemper, assignor to Palm, Incorporated
15 September 2009 (Class 455/575.4); filed in Germany 12 December 2002

When this sliding-case phone is open as shown, the rear surface 4 of display section 2 functions as a bending wave loudspeaker. That is it, folks.



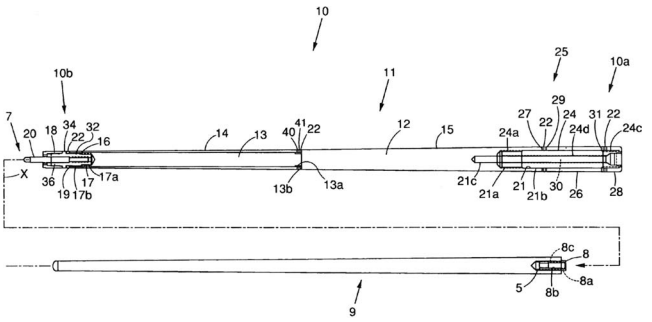
Watch this space for the next big audio advancement from Palm Inc.—GLA

7,559,847

43.40.Kd CUE STICK

Paul D. Costain and Karim Belhaj, assignors to Clawson Custom Cues, Incorporated
14 July 2009 (Class 473/44); filed 17 July 2007

Vibration damping inserts 16 and 24 are placed in the proximal and distal ends, respectively, of butt portion 11 of cue stick 10, which can be, but are not limited to, linen reinforced phenolic material, such as NEMA L Type



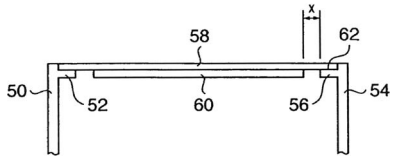
FBI or NEMA LE Type FEI, and are said to damp the vibrations in the stick when the distal end of shaft portion 9 strikes a billiard ball.—NAS

7,568,984

43.40.Tm HOLLOW GOLF CLUB HEAD

Hideo Matsunaga, assignor to Bridgestone Sports Company, Limited
4 August 2009 (Class 473/332); filed in Japan 23 August 2005

The patent states that to increase the launch angle of a golf ball when struck by a hollow golf club head, which are known as drivers, the ratio of rigidity of the sole portion of the club head to the side portion should be



1:0.2–0.6, the ratio of the average thickness of the sole portion to the crown portion should be 1:0.5–0.7, and the ratio of Young's modulus of the materials used in the sole portion and the crown portion should be 1:0.5–0.8. Two graphs showing several values for the specified ratios versus launch angle and initial ball speed are presented. However, when the crown portion rigidity is reduced, the elastic deformation is increased relative to "ordinary" club heads, so the crown portion "rings" and "excellent ball feel cannot sometimes be obtained." To cure this deficiency, while retaining the benefits of the preferred ratios, vibration absorbing layer 60 is bonded or otherwise affixed to crown portion 58.—NAS

7,591,209

43.40.Tm VIBRATION SUPPRESSING CUTTING TOOL

Daisuke Murakami *et al.*, assignors to Sumitomo Electric Hardmetal Corporation
22 September 2009 (Class 82/158); filed in Japan 26 November 2003

Vibration suppression is accomplished by means of a mass that is located in a cavity in the tool and that essentially rattles as the tool vibrates. The mass is made of a material whose density is greater than that of the tool's material. Clearances of 0.01–0.5 mm are suggested between the surfaces of the mass and the sides of the cavity enclosing it.—EEU

7,592,901

43.40.Tm INPUT DEVICE

Hidetaka Furusho, assignor to Alps Electric Company, Limited
22 September 2009 (Class 340/407.1); filed in Japan 25 August 2004

In order to provide tactile feedback to a touch-screen device's user when he touches a point on the screen, the back of the screen is provided with a planar electrode array. The electrodes are arranged in stripes of alternating polarity so that localized areas can be made to vibrate by application of a voltage via an external driving circuit.—EEU

7,592,801

43.40.Yq METHODS AND APPARATUS FOR VIBRATION DETECTION

James M. Bailey *et al.*, assignors to Allegro Microsystems, Incorporated
22 September 2009 (Class 324/207.25); filed 21 March 2008

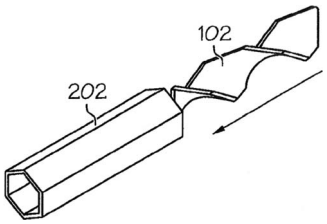
This patent pertains to means for differentiating between rotational and lateral vibration motions of rotating ferrous objects, such as gears. It takes note of the fact that a magnetic proximity sensor can only detect the change in the magnetic field as an object, such as a gear tooth, and moves relative to the sensor; the sensor cannot differentiate between the gear vibrating in rotation relative to the sensor and oscillatory changes in the air gap between the sensor and gear. The method of this patent overcomes this problem by using a plurality of magnetic proximity sensors, rotation detectors, and rotation-agreement processors.—EEU

7,562,742

43.50.Gf AIR EXHAUST/INLET SOUND ATTENUATION MECHANISM

Don A. Gilliland *et al.*, assignors to International Business Machines Corporation
21 July 2009 (Class 181/280); filed 13 June 2008

Inner member 102 has a 360 deg twist and is shaped so that it fits, snugly, in outer member 202. As such the axial twist provides a labyrinth path for both air flow and electromagnetic waves. By suitable choice of



materials, which should have suitable sound damping and electromagnetic attenuating characteristics, for the inner and outer members, combined with the hex-packing design of the other sleeve, a method and system are disclosed that allow forced air cooling of an electronic device as well as electromagnetic interference shielding through the air vent.—NAS

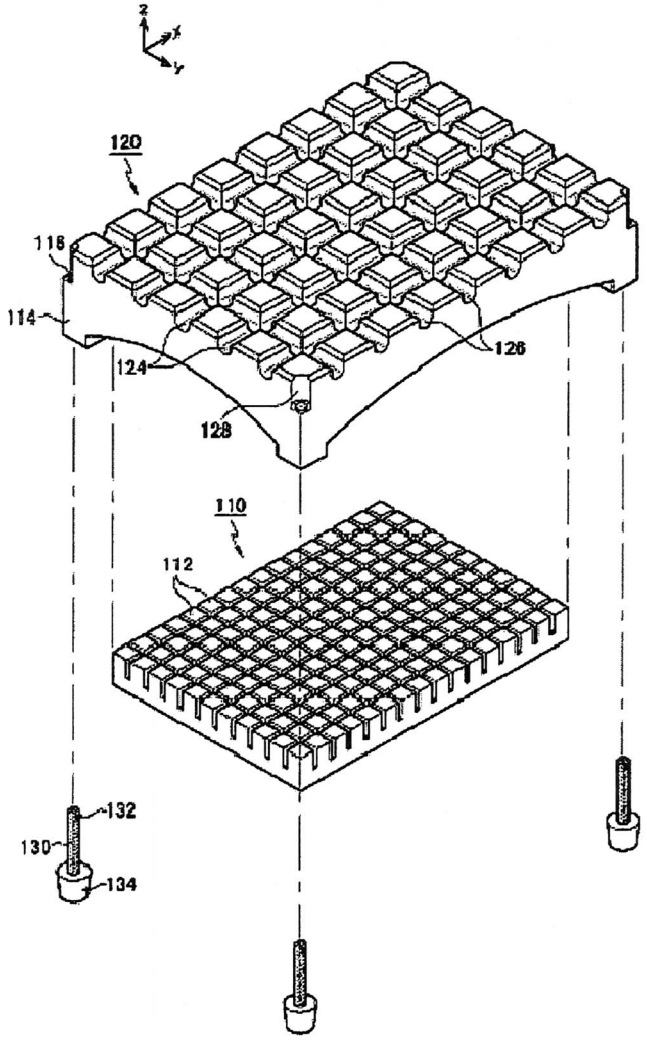
7,594,365

43.55.Ti HEATING FLOOR SYSTEM

Se-Chang Kang *et al.*, assignors to LG Chem, Limited
29 September 2009 (Class 52/220.2); filed in Republic of Korea 15 September 2005

The patent describes a polypropylene panel that sits on four corner legs and supports radiant heating pipes in the grooves on the upper side of the panel. The legs reduce point of contact with the floor structure below and thus reduce sound transfer. The protrusions under the panel (defined by the

grooves on the upper side) interlock with a thermal insulating panel.



The system will save time and money because of the ease of installation.—CJR

7,586,239

43.58.Kr MEMS VIBRATING STRUCTURE USING A SINGLE-CRYSTAL PIEZOELECTRIC THIN FILM LAYER

Sheng-Shian Li *et al.*, assignors to RF Micro Devices, Incorporated
8 September 2009 (Class 310/328.02); filed 6 June 2008

The authors describe a micro-electronic mechanical structure that incorporates a thin film of single-crystal piezoelectric material such as lithium niobate or lithium tantalate into a lateral or thickness mode resonator structure in silicon. The silicon is micromachined so as to be suspended by supporting arms only, confining the resonances to an area whose lateral dimensions can be accurately defined and controlled by lithographic techniques. The resulting layered structure has high mechanical Q, it is claimed, because the vibrations are segregated into the suspended volume. This device is useful for the rf electronics industry, where the filters used have frequencies high enough to benefit from the small dimensions of these devices. The construction principles are similar to many other film bulk acoustic resonators, however.—JAH

7,587,052

43.66.Ts HEARING AID AND METHOD FOR ADJUSTING A HEARING AID

Thomas Kasztelan, assignor to Siemens Audiologische Technik GmbH
8 September 2009 (Class 381/60); filed in Germany 21 May 2004

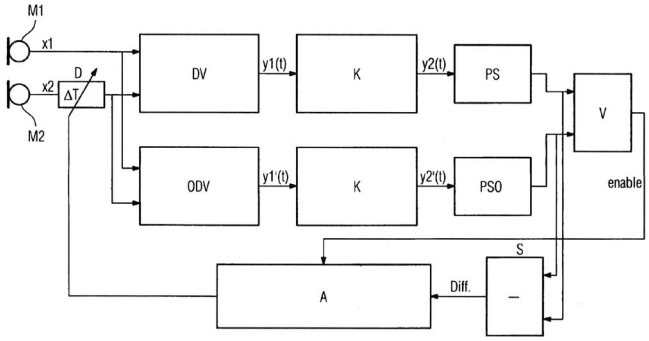
Measured frequency response data obtained while the hearing aid is worn in the user's ear are stored in the hearing aid and utilized to provide faster and more accurate fitting adjustments than would be achieved using average frequency response data stored in the fitting software.—DAP

7,587,058

43.66.Ts METHOD AND DEVICE FOR MATCHING THE PHASES OF MICROPHONE SIGNALS OF A DIRECTIONAL MICROPHONE OF A HEARING AID

Eghart Fischer and Henning Puder, assignors to Siemens Audiologische Technik GmbH
8 September 2009 (Class 381/313); filed in Germany 5 March 2004

To match the phase of two microphone outputs in a directional system, the omnidirectional and directional signal level outputs are matched without



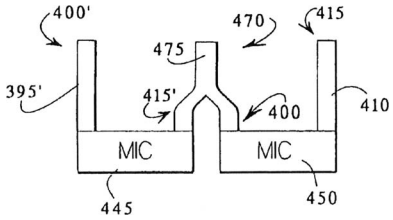
knowing the position of the acoustic source generating the test signals by adjusting a delay applied to one of the microphone outputs. The delay is adjusted only if the directional signal level is higher than the omnidirectional signal level.—DAP

7,590,253

43.66.Ts HEARING AID HAVING SWITCHABLE FIRST AND SECOND ORDER DIRECTIONAL RESPONSES

Mead C. Killion, assignors to Etymotic Research, Incorporated
15 September 2009 (Class 381/313); filed 25 October 2006

A custom hearing aid comprises at least two microphone housings and a faceplate having at least two sound passages acoustically coupling sound



from one or more of the sound openings to the microphone housings. Diffraction scoops with wind screens are placed over the sound openings. The hearing aid switches between an omnidirectional response and a directional response equalized to the omnidirectional response. The switching can be

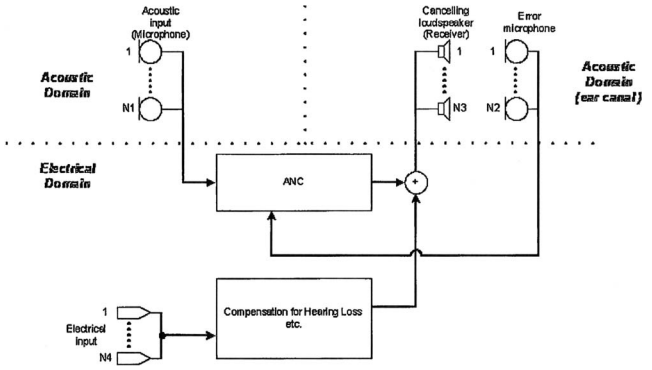
either done manually via a switch or automatically depending on the ambient noise level.—DAP

7,590,254

43.66.Ts HEARING AID WITH ACTIVE NOISE CANCELING

Jes Olsen, assignor to Oticon A/S
15 September 2009 (Class 381/328); filed in Denmark 26 November 2003

One commonly employed method of reducing the sensation of occlusion in hearing aid fittings is to leave the ear canal partially open, which also allows ambient sounds into the ear canal by bypassing the hearing aid signal processing. To reduce such unwanted ambient sounds when a wireless signal



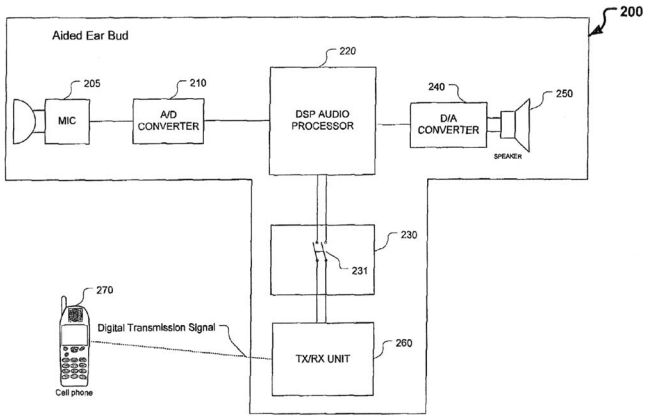
or direct audio input signal is received, the output of an internal microphone picking up sound in the ear canal is used as an error signal, in conjunction with the ambient microphone pickup(s), to generate a compensation signal that partially cancels the ambient sound in the ear canal.—DAP

7,593,537

43.66.Ts AIDED EAR BUD

Mark Enzmann, assignor to AT&T Mobility II LLC
22 September 2009 (Class 381/315); filed 7 June 2007

A hearing aid switches between two operating modes via commands received from a mobile device when answering an incoming call or initiating an outgoing call. One operating mode is normal hearing aid functionality. In the other mode, the hearing aid microphone picks up the user's voice and transmits it to the mobile device and receives an audio signal from the mobile device representing call content. This received signal is mixed with



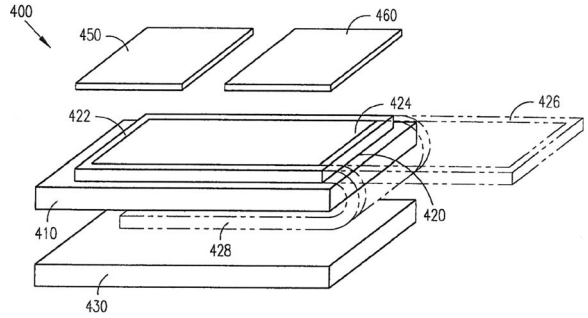
the hearing aid microphone signal, depending on the ambient noise level, and modified by a frequency-amplification map stored in the hearing aid according to the listening needs of the user. Transmissions between the mobile device and hearing aid may be wireless.—DAP

7,593,538

43.66.Ts ANTENNAS FOR HEARING AIDS

Beau Jay Polinske, assignor to Starkey Laboratories, Incorporated
22 September 2009 (Class 381/322); filed 17 February 2006

A flexible antenna, comprising at least one flexible metal trace connected to communication electronics in a hearing aid, is adapted for assembly with a hybrid circuit. The antenna may have several planar coils that



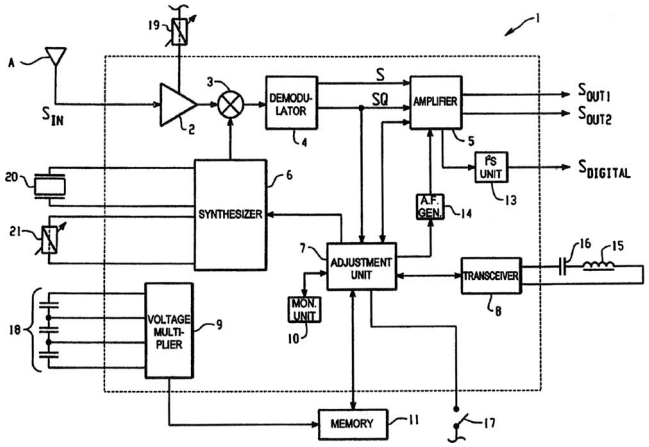
may be connected to form a helical coil.—DAP

7,596,237

43.66.Ts METHOD FOR CONTROLLING A TRANSMISSION SYSTEM, APPLICATION OF THE METHOD, A TRANSMISSION SYSTEM, A RECEIVER AND A HEARING AID

Jean-Claude Constantin, assignor to Phonak AG
29 September 2009 (Class 381/314); filed 18 September 2000

Audio signals are transmitted wirelessly on independent transmission channels to a number of hearing aids, which have been configured from parameters transmitted wirelessly on independent transmission channels. Each audio signal is transmitted at a predefined carrier frequency.—DAP



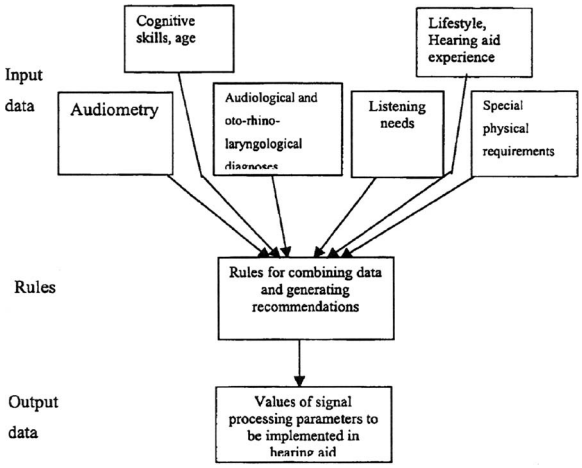
7,599,499

43.66.Ts METHOD FOR FITTING A HEARING AID TO THE NEEDS OF A HEARING AID USER AND ASSISTIVE TOOL FOR USE WHEN FITTING A HEARING AID TO A HEARING AID USER

Graham Naylor, assignor to Oticon A/S
6 October 2009 (Class 381/60); filed in Denmark 28 September 2001

Initial hearing aid signal processing parameters are derived from the

wearer's audiological data and level of cognitive functionality. Cognitive functionality is determined by the wearer's age, medical records, present or previous occupation, or a cognitive performance test or conversation with



the wearer or persons associated with the wearer. Optionally, the wearer's lifestyle, sound environments, and previous hearing aid experience may also be taken into consideration.—DAP

7,599,507

43.66.Ts HEARING AID AND A METHOD FOR ENHANCING SPEECH INTELLIGIBILITY

Martin Hansen, assignor to Widex A/S
6 October 2009 (Class 381/317); filed 12 January 2005

After the hearing aid or an external processor analyzes the sound environment surrounding the hearing aid wearer, an estimate of speech intelligibility is made by the hearing aid or external processor from the initial hearing aid gain vector values, ambient speech levels, and noise levels in a set of frequency bands. The gain vector values in a number of frequency bands are modified automatically to maximize speech intelligibility. The gain vector values may also be adjusted to maintain a calculated loudness below a loudness upper limit.—DAP

7,239,999

43.72.−p SPEED CONTROL PLAYBACK OF PARAMETRIC SPEECH ENCODED DIGITAL AUDIO

Changwon D. Rhee, assignor to Intel Corporation
3 July 2007 (Class 704/211); filed 23 July 2002

This patent presents a simple means for solving a simple problem with pitch-corrected speed control applied to coded speech playback. Various schemes are described for properly equalizing the number of audio samples contained in speech frames received from a decoder (such as linear predictive), with the number of audio samples that are to be used by a prior-art type of playback speed controller.—SAF

7,246,058

43.72.Ar DETECTING VOICED AND UNVOICED SPEECH USING BOTH ACOUSTIC AND NONACOUSTIC SENSORS

Gregory C. Burnett, assignor to Aliph, Incorporated
17 July 2007 (Class 704/226); filed 30 May 2002

This patent presents methods for detecting voiced speech segments with the aid of a general electromagnetic movement sensor that has been patented separately. This latter sends out weak radio waves that reflect off the target. The reflected waves are combined with the transmitted waves, and the combination is checked for phase shifts, which are presumably caused by movement of the target, such as a speaker's neck or throat tissue.—SAF

7,236,922

43.72.Fx SPEECH RECOGNITION WITH FEEDBACK FROM NATURAL LANGUAGE PROCESSING FOR ADAPTATION OF ACOUSTIC MODEL

Hitoshi Honda *et al.*, assignors to Sony Corporation
26 June 2007 (Class 704/2); filed in Japan 30 September 1999

This patent presents a dialog system with speech recognition and synthesis that all seems to be prior art, with the exception of one innovation. A scheme is included for performing speaker adaptation on-line, which is again according to prior art, except for the idea to use information from the dialog system to assign "zones" in the dialog that are likely to provide good-quality adaptation results. The dialog structure is used to detect erroneous recognition, and these zones of the dialog are then omitted from the speech that will be used to perform adaptation.—SAF

7,236,931

43.72.Fx SYSTEMS AND METHODS FOR AUTOMATIC ACOUSTIC SPEAKER ADAPTATION IN COMPUTER-ASSISTED TRANSCRIPTION SYSTEMS

Chuang He and Jianxiong Wu, assignors to USB AG, Stamford Branch
26 June 2007 (Class 704/235); filed 28 April 2003

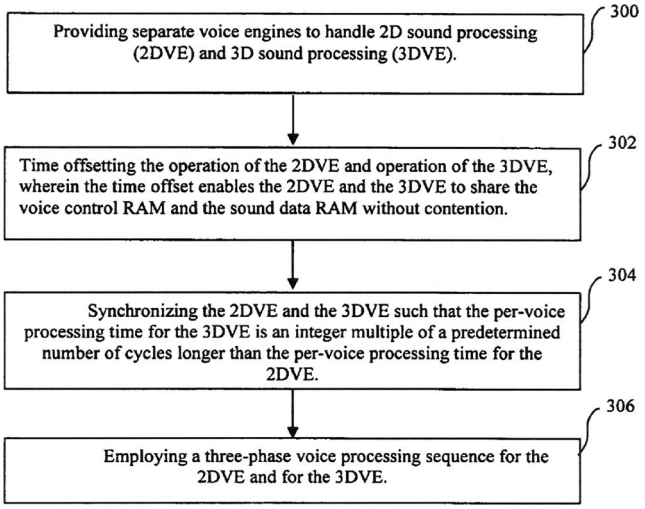
A transcription system is described which performs speaker adaptation by making use of "semi-literal" transcripts that were created at an initial stage by speech recognition using the speaker-independent acoustic models. Various means are put forth for creating maximally correct semi-literal transcripts prior to adaptation.—SAF

7,587,310

43.72.Gy SOUND PROCESSOR ARCHITECTURE USING SINGLE PORT MEMORY UNIT

David H. Lin, assignor to LSI Corporation
8 September 2009 (Class 704/200); filed 30 August 2004

For power and cost savings during at least one of three processing



phases, the two-dimensional (2D) and three-dimensional (3D) voice engines

in a sound processor are synchronized and share at least one single port memory by time-multiplexing without contention. The mixer in the sound processor has separate buffers for the 2D voices, 3D voices, and reverberation.—DAP

7,240,005

43.72.Ja METHOD OF CONTROLLING HIGH-SPEED READING IN A TEXT-TO-SPEECH CONVERSION SYSTEM

Keiichi Chihara, assignor to Oki Electric Industry Company, Limited
3 July 2007 (Class 704/267); filed in Japan 26 June 2001

Speech synthesis systems are sometimes used to perform text read-out at a high speech rate. The patent notes that standard speech synthesis prosody is totally inappropriate for extremely high speech rates, and a scheme is proposed for altering the prosody in various ways to render speech more intelligible at high speed.—SAF

7,249,020

43.72.Ja VOICE SYNTHESIZING METHOD USING INDEPENDENT SAMPLING FREQUENCIES AND APPARATUS THEREFOR

Reishi Kondo, assignor to NEC Corporation
24 July 2007 (Class 704/258); filed in Japan 18 April 2001

This incredibly mundane and obvious patent describes a means of selecting a lower sampling rate for voiced speech segments and a higher sampling rate for unvoiced speech segments stored in a speech synthesis sound database. The objective is to reduce storage requirements.—SAF

7,588,538

43.80.Vj ULTRASONIC DIAGNOSTIC EQUIPMENT AND IMAGE PROCESSING APPARATUS

Yoichi Ogasawara *et al.*, assignors to Kabushiki Kaisha Toshiba
15 September 2009 (Class 600/454); filed in Japan 11 September 2003

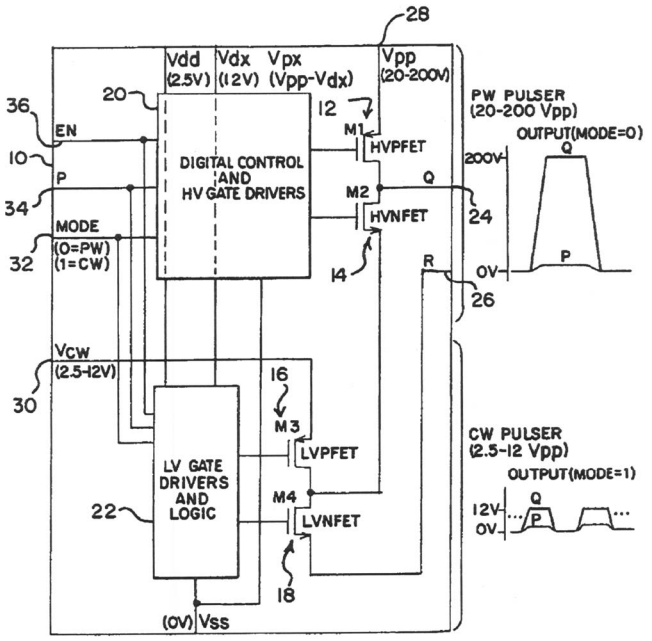
After a contrast agent comprised of gas bubbles is injected into a subject, b-scan imaging is performed using a high acoustic intensity to acquire the maximum echo intensities and to break the bubbles. Subsequent b-scan imaging is performed at different times by using low or moderate acoustic intensities that do not break contrast medium bubbles. Data obtained from the scans are analyzed and displayed.—RCW

7,588,539

43.80.Vj INTEGRATED LOW-POWER PW/CW TRANSMITTER

David A. Petersen, assignor to Siemens Medical Solutions USA, Incorporated
15 September 2009 (Class 600/459); filed 19 January 2005

High-voltage and low-voltage switches are integrated on a semiconductor chip. The high-voltage switches are used for pulse-wave operation and the low-voltage switches are used for continuous-wave operation. Both



the pulse and continuous waveforms come from a common output.—RCW

7,588,540

43.80.Vj ULTRASONIC PROBE FOR SCANNING A VOLUME

An Nguyen-Dinh *et al.*, assignors to Vermon
15 September 2009 (Class 600/459); filed 8 April 2005

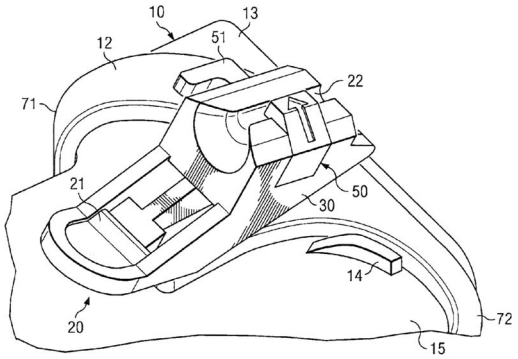
This probe includes a flexible membrane seal. The membrane permits immersion in a coupling fluid through which the probe moves to scan planes in succession. Portions of the membrane also reduce parasitic drag forces produced during movement of the probe.—RCW

7,588,541

43.80.Vj METHOD AND SYSTEM FOR POSITIONING A MEDICAL DEVICE AT ONE OR MORE ANGLES RELATIVE TO AN IMAGING PROBE

Jared Floyd and Steven Bunce, assignors to SonoSite, Incorporated
15 September 2009 (Class 600/461); filed 28 January 2004

A needle guide is attached to the end of an ultrasound imaging probe. The user positions the needle by looking at an ultrasound image produced



with the probe containing the needle. Both the needle and the guide can be released from the probe.—RCW

7,591,786

43.80.Vj DOCK FOR CONNECTING PERIPHERAL DEVICES TO A MODULAR DIAGNOSTIC ULTRASOUND APPARATUS

Randy T. Holmberg *et al.*, assignors to SonoSite, Incorporated
22 September 2009 (Class 600/437); filed 31 January 2003

This apparatus contains a system electronics module, input-output ports, and docking stations that can be released from the system module.—RCW

7,591,787

43.80.Vj METHOD FOR REMOVING DOPPLER ANGLE AMBIGUITY

Piero Tortoli, Firenze, Italy
22 September 2009 (Class 600/454); filed 15 September 2005

An ultrasound beam is initially oriented in a direction essentially perpendicular to the direction of blood flow, and a Doppler spectrum of blood echoes is obtained. The beam is then re-oriented so the Doppler spectrum of blood echoes is essentially symmetric around the zero frequency, and the Doppler spectrum is again determined. The rate of blood flow is calculated using the angle between the ultrasound beams and the Doppler frequency of the echoes obtained from the second ultrasound beam.—RCW

7,591,788

43.80.Vj ADAPTIVE CONTRAST AGENT MEDICAL IMAGING

Patrick J. Phillips and Ismayil Guracar, assignors to Siemens Medical Solutions USA, Incorporated
22 September 2009 (Class 600/458); filed 19 August 2003

A contrast agent imaging mode is dynamically determined and optimized by using a selected transmit level or real-time measurement of the contrast agent signal. An interface allows this imaging mode to be used without a need to switch between multiple modes or contrast agent detection techniques and associated contrast agent imaging parameters. This permits an imaging session without a user switching the mode of system operation.—RCW

7,593,554

43.80.Vj SYSTEM AND METHOD FOR COMPARING ULTRASOUND IMAGES CORRESPONDING TO TWO USER-SELECTED DATA POINTS

Edward A. Miller *et al.*, assignors to Koninklijke Philips Electronics N.V.
22 September 2009 (Class 382/128); filed 3 October 2002

A graphical user interface displays curves that provide data about a sequence of ultrasound images. By selecting a point on a curve, a corresponding ultrasound image is displayed. By selecting another point on the same or another curve, another corresponding ultrasound image is displayed. The two ultrasound frames can then be compared.—RCW

7,597,663

43.80.Vj ADJUNCTIVE ULTRASOUND PROCESSING AND DISPLAY FOR BREAST CANCER SCREENING

Shih-Ping Wang *et al.*, assignors to U-Systems, Incorporated
6 October 2009 (Class 600/437); filed 27 November 2002

Ultrasound scans spanning a volume are displayed. A user can navigate among images showing a thick cross section, a selected enlarged thick cross section, and a planar image. The contrast in the thick cross-sectional images can be inverted prior to display so that bright structures in an x-ray mammogram are also bright in the ultrasound image.—RCW

7,597,664

43.80.Vj MEASURING TRANSDUCER MOVEMENT METHODS AND SYSTEMS FOR MULTI-DIMENSIONAL ULTRASOUND IMAGING

Paul H. Jones *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
6 October 2009 (Class 600/437); filed 30 October 2006

An optical sensor in an ultrasound transducer measures motion along the skin in a way similar to the way that a mouse in a personal computer measures movement. The transducer can be tilted at an angle toward or away from the direction in which the transducer is translated. The distance of travel is determined by angle-correcting the movement that is measured.—RCW

Session 1aAAa

Architectural Acoustics Noise, and INCE: Acoustics of Green Buildings

Brandon D. Tinianov, Cochair
Serious Materials, 1250 Elko Dr., Sunnyvale, CA 94089-2213

David M. Sykes, Cochair
Remington Partners, Inc., 23 Buckingham St., Cambridge, MA 02138

Ralph T. Muehleisen, Cochair
Illinois Inst. of Technology, Civil and Architectural Engineering, 3201 S. Dearborn, Chicago, IL 60616

Chair's Introduction—8:00

Invited Papers

8:05

1aAAa1. US Green Building Council's role as a change agent in the North American construction market. Brandon D. Tinianov (Serious Mater., 1250 Elko Dr., Sunnyvale, CA 94089, btinianov@seriousmaterials.com)

After observing market reaction to the U.S. Green Building Councils (USGBC) Leadership in Energy and Environmental Design (LEED), one could conclude that it may be the most powerful voluntary market transformation tool encountered in the modern building industry. The LEED voluntary system is not only entering a very diverse array of markets but stimulating changes in product design, professional practices, and educational curricula. It is also prompting debate in various circles another healthy sign of market transformation. This presentation will provide an overview of who USGBC represents, how they view green buildings, and the specifics of how this market transformation takes place.

8:20

1aAAa2. A review of the new Leadership in Energy and Environmental Design v3 green building rating system. Ralph T. Muehleisen (Civil, Arch. and Env. Eng., Ill. Inst. of Tech., Chicago, IL 60616, muehleisen@iit.edu)

Leadership in Energy and Environmental Design (LEED) has become the *de facto* standard green building rating system in the US. On April 27, 2009 the US Green Building Council launched LEED v3 which is a major change to the previous version, becoming more flexible and prioritizing energy efficiency and greenhouse gas emissions. In particular, the prerequisites and credits amongst the various "flavors" of LEED have been harmonized for increased consistency, credit weights have been adjusted to better reflect impact on human and environmental health concerns, and credit weightings now reflect regional environmental issues. Acoustical consultants and engineers need to be aware of and understand these changes because they need to work with architects and other engineers who are designing buildings to meet the new LEED requirements. In this presentation, LEED v3 and the current implementation, LEED 2009, are reviewed.

8:35

1aAAa3. Architectural acoustics: Emerging opportunities require new materials and solutions. David M. Sykes (Remington Partners LLC, 23 Buckingham St., Cambridge, MA 02138, dsykes@healthcareacoustics.org), William Cavanaugh (Cavanaugh Tocci Assoc., Sudbury, MA 01776), Mandy Kachur (Acoust. by Design, Ann Arbor, MI), Erica Ryherd (Georgia Inst. of Technol., Atlanta, GA), and Alexis Kurtz (Arup Acoust., New York, NY)

The landscape is changing. New 2010 code-level acoustical guidelines for healthcare facilities and new 2009 LEED environmental quality criteria for acoustics in green buildings (LEED) create demand for acoustical solutions that stretch the boundaries of current practice. In many cases, these needs cannot be met by traditional methods, materials, technologies, or best practices in design & construction. Unfortunately, technical development has been constrained for 3 decades by absence of government support for R&D and by litigation against many manufacturers of acoustical materials. As a result, serious re-thinking is needed by materials manufacturers, technology developers, solutions providers, research scientists, and consultants. This paper gives an overview of these emerging needs and outlines new ways of thinking about architectural acoustics. Topics include (1) new requirements (renewable sources, VOCs, recyclability, lightweight construction, infection control/cleanability, and daylight transmission); (2) materials (re-purposed, adapted, and new); (3) construction method & procedures (pre-fabricated solutions, & new sources thereof, and the changing role of value engineering); (4) measurement & monitoring method & equipment (MEMS, wireless sensors, smart-phone applications, and cloud computing); and (5) code & rating systems (including post-construction performance inspection & certification procedures and the need for periodic monitoring & re-certification).

8:50

1aAAa4. Unintended consequences: Acoustical outcomes in LEED (Leadership in Energy and Environmental Design) certified offices. Kevin Powell (Office of Appl. Sci., GSA Public Bldgs. Service, 555 Battery St., Rm. 518, San Francisco, CA 94111, kevin.powell@gsa.gov)

Recent research conducted by the US General Services Administration and the Center for the Built Environment at UC Berkeley disclosed an anomaly about "green building" interiors that is an unintended consequence of the current LEED rating system. The anomaly is this: occupants of offices that have been upgraded to achieve LEED certification express high levels of satisfaction with the "indoor environmental quality" of almost all attributes of these spaces; but there is one exception where occupant scores are lower than before. Their increased dissatisfaction has to do with acoustics and privacy. What is it about LEED-rated offices that cause increased dissatisfaction? Why are occupants uncomfortable with the "acoustical comfort" of the new LEED-certified offices? Is it because there are more reflective, hard surfaces—such as exposed ceilings—causing noise to reverberate? Or because HVAC systems have become quieter? Or because partition heights have been lowered to allow better airflow and more natural light, resulting in less privacy? Since increased occupant dissatisfaction may result in reduced productivity or increased absenteeism, the GSA's Office of Research has been examining this issue and report on its findings and conclusions.

9:05

1aAAa5. Understanding and specifying green noise control materials. Brandon D. Tinianov (Serious Mater., 1250 Elko Dr., Sunnyvale, CA 94089, BTinianov@seriousmaterials.com)

Specification of green acoustical materials is difficult to implement from the acoustical expert's perspective. A contributing factor is the accepted definition of "green" that is poorly understood by the acoustical community. The widely accepted interpretation of the term does not refer exclusively to building materials that have a high recycled material content or are made without the use of toxic chemicals but extends to the entire project. Overall building design implications extend beyond material content to water consumption, energy use, and to the health and wellness of the occupants. These larger considerations are reflected in the most prevalent green rating system, the LEED rating system as established by the United States Green Building Council (USGBC). Within the framework of a larger, holistic building system, acoustical material can be considered green on three levels: (a) its contribution to the health and comfort of the building occupants, (b) environmentally preferable nature of its components or manufacturing process, and (c) its secondary performance features that have a positive impact on other building systems. Acoustical experts must therefore review the "green-ness" of sound control materials in this same framework. A review of such an approach and a review of existing green acoustical products will be presented.

9:20

1aAAa6. Acoustics and the leadership in energy and environmental design rating systems. Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

The leadership in energy and environmental design (LEED) rating systems address acoustics explicitly in but a few of their many credits. More profound are the implications of LEED-driven design on a building acoustics, typically in pursuit of LEED credits where no explicit reference to acoustics, noise, or acoustical design exists. Using project case studies as a guide, this presentation will outline some of the conflicts as well as synergies between "best practice" acoustical design and design decisions made on the basis of achieving certain LEED credits. A range of example credits will be presented along with the ramifications of such credits on building acoustics, including the effects of sustainable site credits on environmental noise isolation, energy and atmosphere credits on background noise levels, materials and resources credits on acoustical material specification, and indoor environmental quality credits on sound isolation and other acoustical design elements.

9:35

1aAAa7. Acoustics and the Leadership in Energy and Environmental Design (LEED) certification—Why education and healthcare came first. Kurt Rockstroh (Steffian Bradley Architects, 100 Summer St., Boston, MA 02110-2106, kurtr@steffian.com), David Sykes (Remington Partners, Cambridge, MA 02138), William Cavanaugh, and Gregory Tocci (Cavanaugh Tocci Assoc., Sudbury, MA)

Three decades of Washington neglecting the noise problem in America resulted in general ignorance of and data about human health impacts. This is true with politicians, agency heads, foundations, doctors, educators, architects and builders—whether they "advocate green policies" or not. So when LEED redefined the term "environmental quality" to include acoustics, they focused on "populations at risk"—children (schools), and sick people (healthcare facilities where scientific research on human subjects is routine). Meanwhile, healthcare professionals became attuned to noise and privacy from HIPAA and the patient-centered-care movement. In 2005, the Facility Guidelines Institute, which each re-issues the 60-year-old Guidelines for Design and Construction of Healthcare Facilities every 4 years resolved to address this problem. We commissioned TC-AA.NS.SC to develop comprehensive acoustical guidelines which were completed in 2006 and launched this January. Once drafted, the acoustical guidelines were quickly piloted for LEED by Green Guide for Healthcare, and then the LEED Environmental Quality Technical Advisory Group adopted them as the Reference Standard for LEED HC. From HC it will find broader acceptance in other LEED programs. This process ensures that acoustics re-enters public policy supported by independent, transdisciplinary research on human health impacts and by multiple rounds of peer review in healthcare.

9:50

1aAAa8. Acoustical materials for a green world: The sustainable design transformation of the architectural acoustics industry. Mandy Kachur (Acoust. by Design, Inc., 303 Detroit St., Ste. 304, Ann Arbor, MI 48104, mandy@acousticsbydesign.com), Daniel Bruck, and Jeanette Hesedahl (BRC Acoust. & Technol. Consulting, Seattle, WA 98134)

The focus on sustainable and healthy building materials has intensified over the last few years. This is true particularly in healthcare, educational, and environmentally friendly or "green," facilities, where an increasing awareness of acoustics has resulted in a higher demand for innovative products. As the healthcare industry pushes for greater patient comfort, increased speech privacy, and green design, new acoustical products must satisfy multiple requirements. In educational facilities, the U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) for Schools rating system promotes the use of green products, which extends to acoustical materials. Additionally, some educational facility owners are hesitant to allow traditional types of sound absorbers such as fiberglass duct liner for fear of health issues. Consequently, acoustical materials manufacturers have responded with new and repurposed materials to suit these needs. Because many of these materials have been recently introduced, a significant amount of project time can be expended on research to select an appropriate material to suit the new requirements. In an attempt to streamline the selection process, this overview will identify a number of these new and repurposed materials and provide feedback on their use from architects, engineers, and owners.

10:05—10:15 Break

10:15

1aAAa9. Aerogels: A "Green" thermo-acoustic insulation material with nanoscale properties. James Satterwhite (Cabot Corp., 157 Concord Rd., Billerica, MA 01821, james_satterwhite@cabot-corp.com)

Aerogels are a well-known class of thermal insulation derived from nanoscience that has "green" benefits including translucence, thinness, hydrophobicity, light weight, and flexibility. Recently, the acoustical properties of aerogels have been characterized. Aerogels are currently available in building materials like skylights and exterior glazing, fabrics-based roofing membranes, and flexible blankets for insulating underwater pipelines and building walls. In 2008–2009, laboratory testing and field research began on the acoustical properties of thin profile (2–8-mm) architectural "tensile membrane" fabrics incorporating silica aerogel granules. Data from a tension structure in Canada—where aerogel-enhanced fabric was used to block aircraft noise—exhibited excellent acoustic absorption and acoustic impedance matching properties compared to insulators of comparable thickness. The material increased transmission loss of exterior to interior noise and also reduced indoor reverberation. In the same period, US field tests demonstrated an aerogel blanket material as a surface treatment in open offices to reduce broadband reverberation, resulting in increased speech intelligibility and privacy and enhanced acoustical comfort. These acoustical attributes combined with aerogels' thermal value, thin form factor, translucence, hydrophobicity, light weight, and absence of VOCs had led to growing interest in green building applications ranging from aircraft interiors to hospitals.

10:30

1aAAa10. Materials science: A "Green" acoustical wall treatment for controlling room reverberation in healthcare. Francis Babineau (Johns Manville, 10100 West Ute Ave., Littleton, CO 80127, francis.babineau@jm.com)

In healthcare, green criteria must be coupled with a higher standard of stringent infection control requirements. This study was conducted to explore an alternate method for reducing room reverberation in a healthcare setting. Reverberation time (RT) plays an important role in noise levels and speech privacy which are key factors in healthcare. However, RT is difficult to control in healthcare facilities due to strict durability and infection control requirements. In this study of a private patient room in a hospital, instead of improving the ceiling to reduce reverberation, the walls were treated with a new, thin, acoustically absorptive, but paintable, wall material. The wall treatment was a new-to-market acoustical finish with absorption coefficients ranging from 0.05–0.6. To achieve reverberation control, several walls of the room were covered with the treatment. Reverberation times were measured before and after installation, and results were consistent with predicted values. Despite being a small study, results suggest that a thin, paintable, acoustical wall treatment that meets criteria for infection control and durability can be an effective method for controlling reverberation in healthcare facilities, thereby creating a more comfortable environment for patients and staff while giving architects increased design flexibility.

10:45

1aAAa11. Going green may require new professional relationships and different solution providers. Bennett Brooks (27 Hartford Turnpike, Vernon, CT 06066, bbrooks@brooks-acoustics.com) and Jason Perrone (BKM Technol. Group, East Hartford, CT 06108)

Change is always welcome but often confusing. One way climate change impacts the acoustics profession is through the sudden, government-driven rush to make buildings green by conforming to Energy Star and the U.S. Green Building Council's LEED rating system (acoustics have only recently become part of the "environmental quality" vocabulary of these systems). Two more ways change impacts acousticians are first the emergence of technologically advanced materials and devices (often from companies you do not know), and second, the emergence of new "solution-providers" with prior contractual relationships with your clients. The presenters represent three entities that have been cooperating to address green-related acoustical issues for large corporate clients in the Northeast USA. The projects described in this presentation have been instructive for all three parties: new materials have been put to use in unexpected ways, independent testing validated the efficacy of novel solutions, and client relationships were managed by a corporation previously unrecognized by most acoustical engineers but that has spent a decade establishing a level of skill in acoustics appreciated by government agencies and leading corporations that have streamlined supply chains and therefore seek to use a limited number of existing supplier relationships to implement change.

11:00

1aAAa12. Electro-acoustic architecture: Is it Green? Roger W. Schwenke (Meyer Sound Labs., 2832 San Pablo Ave, Berkeley, CA 94702, rogers@meyersound.com) and Jason R. Duty (Charles M. Salter Assoc., Inc., San Francisco, CA 94104)

Electro-acoustic architecture systems offer a means of changing the acoustic properties of a room electronically. They are an alternative to physically variable acoustics, where acoustic properties are changed by retractable curtains and doors opening to reverberant chambers, etc. This paper will address the question of whether electro-acoustic systems are a Green alternative to physically variable acoustics. In order to have a reverberation time appropriate for symphonic music, a room must have a large cubic volume and hard heavy surfaces. In contrast, in a room with an electro-acoustic system, the physical reverberation time is desired to be low even at low frequencies. Therefore, the cubic volume can be low and surfaces can be made of a lightweight material over an airspace. This affects the amount of energy needed to deliver materials, construct the building, and operate HVAC and lighting systems. A lower volume building means less pollution and waste during construction, and allows for more open space around the building, and/or higher development density. Electro-acoustic systems can easily be designed as a renovation, thus recycling existing buildings. Case studies will be reviewed demonstrating these issues.

11:15

1aAAa13. Green classroom acoustics: Sound or noise? Glenn E. Sweitzer (Sweitzer LLP, 4504 N. Hereford Dr., Muncie, IN 47304, glenn.sweitzer@gmail.com)

Architectural acoustics is underrepresented in current green building rating systems. While acoustics factors are embodied in green building issues, their scoring potentials remain untapped. An alternative, individual user-oriented basis is tested for scoring architectural acoustics factors. Key architectural acoustics factors embodied in the currently prevailing LEED Green Building Rating System are identified. Selected acoustics factors, and relevant criteria, are studied in the context of four similar school classrooms in one K-5 school. To assess the relative values of these factors, four K-5 teachers are surveyed and then interviewed regarding sound and noise control strategies used in their respective classrooms by academic subject, visual and aural media, and teacher-student and student-student configurations . Reported outcomes are compared with measured sound transmission index (STI) values for selected activity configurations in each classroom. Results suggest that passive sound and noise control in individual classrooms can be used to improve STI by academic subject, presentation media, and teacher and student configurations. Accordingly, scoring for an individual classroom will vary by the passive as well as active noise control means used. Examples are discussed.

11:30—12:00 Panel Discussion

MONDAY MORNING, 19 APRIL 2010 GRAND BALLROOM III/IV, 11:00 A.M. TO 12:00 NOON

Session 1aAAb

Architectural Acoustics: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

William J. Cavanaugh, Cochair
Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776

David Lubman, Cochair
DL Acoustics, 14301 Middletown Ln., Westminster, CA 92683-4514

Chair's Introduction—11:00

Invited Paper

11:05

1aAAb1. Communication behavior of pupils and teachers in highly absorbent classrooms: A pleading for room acoustic comfort in schools. Markus Oberdoerster (Saint-Gobain Ecophon GmbH, Taschenmacher Str. 8, 23566 Lbeck, Germany, markus.oberdoerster@ecophon.de) and Gerhart Tiesler (Univ. of Bremen, 28334 Bremen, Germany)

This lecture refers to an interdisciplinary research project conducted 2001–2005 by acousticians, pedagogues, and occupational scientists at the Bremen University. It investigates the kind of work and communication behavior in elementary school classrooms and the influence that is given by the acoustic environment. Using a database of 175 examined lessons an analysis is made of how different kinds of work (frontal lessons vs differentiated lessons) have an effect on the sound level in the classroom. Parameters are discussed, which can describe classroom acoustics appropriately. Also discussed are how altered room characteristics (e.g., increased absorption, shortened reverberation time, and improved speech intelligibility) affect the sound level in the context of each kind of work. A methodical examination of the database allows an assessment of mean values but also of the detailed teaching phases, as characterized by certain pedagogical factors. The results provide the basis for discussion of stress and work demands of teachers.

Session 1aAB

Animal Bioacoustics: Effects of Anthropogenic Noise on Aquatic Animals I: Acoustic Envrionment and Behavioral Responses

Carl R. Schilt, Chair
Bigleaf Science Services, P.O. Box 225, North Bonneville, WA 98639

Contributed Papers

8:20

1aAB1. An acoustic modeling graphical-user-interface tool for the marine conservationist. Laurel Henderson, Michael Porter, and Christian de Moustier (Heat, Light, and Sound Res., 3366 N. Torrey Pines Court, La Jolla, CA 92037, laurel@hlsresearch.com)

We develop an interactive tool to estimate the total level of anthropogenic and natural noise received by marine animals. This tool will help the marine conservationist assess potential noise-exposure risks. It uses physics-based underwater-acoustic models from the Acoustics Toolbox (http://oalib.hlsresearch.com/). Input choices include ocean area, seasonal sound speed profiles, and wind speed from NOAA and OAML databases. Sound sources can be chosen individually from a panel or by the dozens from a table with built-in representative values of various noise sources, including ships of various sizes, marine construction, navy sonars, and explosives. The user can either study the effects of the built-in noise sources or alter the table with actual noise-sources at known locations. Databases of marine mammal densities can be overlain, or the received acoustic level at chosen points on the map can be saved. We will review the background of the acoustic models used in the GUI and give a short demonstration of the capabilities of this tool.

8:35

1aAB2. Modeling marine mammal sound exposure levels due to ship traffic noise. Christian de Moustier, Michael B. Porter, and Laurel Henderson (HLS Res., Inc., 3366 North Torrey Pines Court, Ste. 310, La Jolla, CA 92037)

Noise due to ship traffic is a pervasive component of the global and local ocean noise from 10 to 500 Hz, but little is known of its effects on marine mammals because such effects depend on the sound exposure levels at a given location. The approach presented here consists of modeling the sound field due to various spatial distributions of noise sources located 5–10 m below the sea surface. We approximate a three-dimensional sound field by computing the sound propagation from each source along a prescribed number of two-dimensional vertical radial slices of a range-dependent environment, which includes bathymetry, seafloor geoacoustic properties, water column sound speed profiles, and sea surface roughness. Propagation results from all sources are incoherently summed to form the overall sound exposure map. By varying the density of ships in a given area we can simulate the effects of shipping lanes and assess how the characteristics of the underlying seafloor alter the results.

8:50

1aAB3. Investigations of marine mammal exposure to sound scattered from the sea surface. Ahmad Abawi, Michael Porter (Heat, Light, and Sound Res., Inc., La Jolla, CA 92037), and Martin Siderius (Portland State Univ., Portland, OR 97201)

Scattering of sound from the rough sea surface results in variations in sound intensity. If the surface height variations are smooth, as they are during a swell, the surface can act as a concave mirror and amplify the sound intensity at its focal region. We have shown both analytically and using numerical modeling that the intensity at the focal region is a function of frequency: for a singly curved surface, the intensity at the focus varies linearly with frequency and for a doubly curved surface, this dependence is quadratic. Scattering of sound waves from a rough surface can also result in a reduction in intensity through dispersion. In this paper, we investigate the possible effects of exposure to sound scattered from the sea surface on marine mammals by computing the scintillation index as a function of range and depth for a variety of ocean environments, including those with surface ducts. We compute the scintillation index using a Monte Carlo technique, where for each surface height realization, we compute the scattered field in the waveguide exactly using the virtual source technique. We compute the Green's functions needed in the virtual source technique using the Scooter propagation model.

Invited Papers

9:05

1aAB4. Odontocete hearing: Effects of noise and measurement during echolocation. Paul E. Nachtigall (Marine Mammal Res. Program, Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI) and Alexander Ya. Supin (Inst. of Ecology and Evolution, Russian Acad. of Sci., Leninsky Prospekt, Moscow, Russia)

Noise that interferes with active echolocation can come from either outside sources or be self-generated. Our work on temporary threshold shifts has shown that either lower levels of anthropogenic noise presented for long time periods or intense sonar pings for short time periods can produce temporary reduction in sensitivity and temporary threshold shifts. Intense anthropogenic sounds intended to disrupt echolocation can also reduce echolocation performance. Our measurements of hearing during echolocation have shown that self-generated intense outgoing signals are managed by the whale's auditory system by (1) the use of forward masking as an automatic gain control system, (2) anatomical (or neurological) damping of outgoing signals, and (3) active manipulation of sensitivity levels of over all hearing. These processes of hearing allow the whale to adjust its hearing during echolocation to maximize the hearing of echoes. Measurements of hearing while a false killer whale actively echolocated, by measuring auditory evoked potentials, showed that the whale heard returns from a small cylinder better than it heard its own outgoing 180-dB clicks that produced the echo returns from the cylinder. Interactions between outside anthropogenic noise and the hearing of self-generated intense sounds during echolocation remain largely unexplored.

9:25

1aAB5. Observing the behavioral response of herring exposed to mid-frequency sonar signals. Nils Olav Handegard, Lise Doksaeter, Olav Rune Godoe (Inst. of Marine Res., NO-5817 Bergen, Norway, nilsolav@imr.no), and Petter H. Kvadsheim (Maritime Systems Div., Norwegian Defence Res. Establishment, NO-3191 Horten, Norway)

There is general concern on how mid-frequency military sonars might affect aquatic animals. Approaches used to investigate possible effects on Norwegian spring spawning herring (*Clupea harengus*) are presented. Experiments were performed in a sheltered fjord area, in the open ocean, and in a net pen. In the fjord area, the behavior of the exposed herring was monitored using an upward-looking bottom-moored echo sounder. For the open ocean experiment, schools of exposed herring were tracked using omnidirectional sonar. In the controlled net-pen experiments, the herring were towed into a fjord basin. The net-pen was equipped with a horizontal pointing and an upward-looking split-beam echo sounder to monitor the behavior, a hydrophone to measure the exposure levels, and a video camera to verify the observed behavior. In all experiments the herring was exposed to relevant naval sonar signals using operational sources mounted on a research vessel or operated from a frigate. For the net-pen experiments, the herring were also exposed to other audible sound sources to investigate the robustness of the methodology to detect reaction patterns. The pros and cons of the different observational approaches are discussed. [Work supported by the Research Council of Norway Grant No. 184705.]

Contributed Paper

9:45

1aAB6. Loggerhead turtles dive in response to airgun sound exposure. Stacy L. DeRuiter (Acoust. and Seismic Service, Ifremer, BP 70, 29280 Plouzané, France) and Kamel Larbi Doukara (Univ. of Oran Es-Senia, BP 1524, El M'Naouer 31000, Oran, Algeria)

Visual observations of the behavior of over 150 loggerhead turtles (*Caretta caretta*) were collected over a period of 2 weeks during a seismic survey in the Mediterranean Sea off Algeria in September and October 2009. All turtles were observed during active operation of an airgun array that had a peak source level of 252 dB *re* 1 μPa. Recordings from three hull-mounted hydrophones allowed concurrent estimation of near-surface airgun array source signatures and sound exposure levels. Of the 53% of turtles that were successfully visually tracked until they had passed more than 100 m behind the seismic vessel, 51% dived at or before their closest point of approach to the airgun array. Among animals that dived, approximately 20% did so immediately following an airgun shot, often showing a startle response that was clearly distinguishable from the predominant basking behavior. Turtle dive probability as a function of distance from the airgun array and sound exposure level will be considered in detail. The observed diving behavior may be interpreted as an avoidance response and may have negative consequences for turtles if it interferes with thermoregulation (basking) or results in inhabitual energy expenditures.

10:00—10:25 Break

Invited Paper

10:25

1aAB7. Changes in vocal behavior of individual North Atlantic right whales in increased noise. Susan E. Parks (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, sep20@psu.edu), Mark Johnson (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Douglas Nowacek (Duke Univ., Beaufort, NC 28516), and Peter Tyack (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

This study investigates the impacts of anthropogenic noise exposure on the vocal behavior of individual North Atlantic right whales, a baleen whale species found in the urban coastal waters off the east coast of the United States and Canada. A non-invasive acoustic recording tag, the Dtag, was used to record the noise levels received by individual whales and the vocalizations they produced in the Bay of Fundy, Canada. These data were used to assess the variability in the received levels (and therefore source level), duration, and frequency content of calls produced by the tagged whale in varying ambient noise conditions. A single stereotyped call type, the 'upcall,' was selected for these measurements. Individual whales producing multiple calls showed increases in received call amplitude and minimum frequency in increasing low-frequency noise conditions. This is one of the first studies to document call intensity changes in baleen whales in response to short-term changes in anthropogenic noise in their environment. This evidence for individual call modification in response to changes in background noise has implications for both descriptive studies of vocal behavior and design for passive acoustic monitoring systems for marine species.

Contributed Papers

10:45

1aAB8. Acoustic compensation to shipping and airgun noise by Mediterranean fin whales (*Balaenoptera physalus*). Manuel Castellote (Natl. Marine Mammal Lab., Alaska Fisheries Sci. Ctr./NOAA, 7600 Sand Point Way, N.E. F/AKC3, Seattle, WA 98115, mcastellote@hotmail.com), Cristopher W. Clark (Cornell Univ., Ithaca, NY 14850), and Marc O. Lammers (The Hawaii Inst. of Marine Biology, Kaneohe, Hawaii 96744)

Seafloor recorders were deployed in the western Mediterranean Sea and adjacent Atlantic waters during 2006–2009 to monitor noise levels and fin whale presence. Acoustic parameters of 20-Hz pulses (pulse duration, bandwidth and intervals, and center and peak frequencies) were compared for areas with different shipping noise levels, different shipping intensities in the Strait of Gibraltar and during seismic airgun events. Statistically significant differences were detected between noise contexts. In general, acoustic parameter values decreased with increased noise levels. In high-noise conditions, 20-Hz pulse duration shortened, bandwidth decreased, and center and peak frequencies decreased. Similar results were obtained in the presence of airgun events, and bearings to singing whales indicated that whales moved away from the airgun source and out of our detection area for a time period that extended well beyond the duration of the airgun activity. This study provides evidence that fin whales modify their acoustic behavior to compensate for increased ambient noise and shows that under some conditions they will leave an area for an extended period. Sensitization and habituation processes may play a role in these behaviors and are discussed in the context of these results.

11:00

1aAB9. Anthropogenic noise and sound production by beluga whales (*Delphinapterus leucas*) in Cook Inlet, Alaska during early August 2007. Cara F. Hotchkin (Ecology IGDP 165 Appl. Res. Lab, The Pennsylvania State Univ., University Park, PA 16802, cfh121@psu.edu.), Susan E. Parks (The Penn State Univ., State College, PA 16804), and Barbara A. Mahoney (Natl. Marine Fisheries Service, Anchorage, AK 99513)

The beluga whale population in Cook Inlet, Ak, has a limited range centered near the Port of Anchorage and was declared endangered in 2008. One potential threat to the belugas recovery is anthropogenic noise, which may disrupt communication and normal behaviors. This project used a boat-based single hydrophone system to evaluate noise levels at several locations in Cook Inlet on 6 days from 2–14 August, 2007. Belugas were encountered twice, at the Port of Anchorage and near the Little Susitna River. Recorded vocalizations were analyzed to develop a provisional description of the whales' repertoire. Call frequencies ranged from 381 Hz to 24 kHz (the limit of the recording system), with most energy at frequencies above 2 kHz. Most vocalizations were similar to call types of other beluga populations. Broadband and 1/3-octave band levels were evaluated for noise from tidal flow and ships at the Port of Anchorage. Vessel noise levels were highest below 0.5 kHz, but frequencies ranged to greater than 8 kHz (distance to source $<$ 200 m). Based on the overlap in frequency between beluga vocalizations and ship noise, anthropogenic sound may mask beluga communication near vessels in Cook Inlet.

11:15

1aAB10. Cook Inlet beluga whale echolocations at the Port of Anchorage expansion. Ana Širović and Lindsey S. Kendall (Alaska Pacific Univ., 4101 Univ. Dr., Anchorage, AK 99508)

Ocean noise levels increase as a result of human activities. As a result, the effects of anthropogenic noise on marine mammals are a growing concern. A genetically distinct population of beluga whales in Cook Inlet was listed as endangered under the Endangered Species Act in 2008. NOAA has identified many potential threats to the Cook Inlet beluga whale population, including coastal development and anthropogenic noise. The Port of Anchorage Marine Terminal Redevelopment Project, currently underway in Upper Cook Inlet, involves several types of water construction activities, including pile driving. However, the effect of pile driving noise on beluga whales is unknown. A passive acoustic array of sonobuoys was deployed during 20 days in August and September 2009 to acoustically detect the presence of beluga whales in the vicinity of in-water pile driving at the Port project site. Beluga whale echolocation clicks were detected on days when the whales were also visually sighted by the on-site visual observers. Whistles and burst pulses were not commonly detected from beluga whales in this area.

11:30

1aAB11. Hearing ability of the lined seahorse (*Hippocampus erectus*). Susan M. Hastings, Meghan R. Kelly, Lisa L. Wu (Thomas Jefferson High School for Sci. and Technol., 6560 Braddock Rd., Alexandria, VA 22312), and Mardi C. Hastings (6422 Crosswoods Dr., Falls Church, VA 22044, mchastings@verizon.net)

The lined seahorse (*Hippocampus erectus*), a member of the family Syngnathidae (pipefishes and seahorses), occurs in coastal waters of the western Atlantic, Gulf of Mexico, and Caribbean Sea. This harmless species is listed as "vulnerable" on the International Union for the Conservation of Nature (IUCN) List of Threatened Species, meaning that it has a high risk of extinction in the medium-term future primarily due to declines in population from targeted catch, bycatch, and degradation of habitat. However, because little is known about hearing in Syngnathids, it is difficult to assess potential effects of anthropogenic sound on these animals. In this study auditory evoked potentials (AEPs) were measured in four *H. erectus* specimens to determine hearing thresholds and bandwidth. Thresholds were determined using the staircase method with exposures to tone bursts between 50 Hz and 21.6 kHz. At low frequencies *H. erectus* had AEP thresholds similar to other bony fishes; however, at frequencies above 2 kHz its auditory sensitivity was similar to that of clupeiform species which detect ultrasound [Mann *et al.*, J. Acoust. Soc. Am. (2001)]. The results of this study provide hearing data that can be used to understand potential effects of human-generated noise on *Hippocampus erectus* and other seahorse species.

Session 1aAO

Acoustical Oceanography Animal Bioacoustics, and Underwater Acoustics: Acoustics in Polar Environments I

Jennifer L. Miksis-Olds, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

David K. Mellinger, Cochair
Oregon State Univ., 2030 S.E. Marine Science Dr., Newport, OR 97365

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Chair's Introduction—9:00

Invited Papers

9:05

1aAO1. The accessible Arctic Ocean. John W. Farrell, III (US Arctic Res. Commission, 4350 N. Fairfax Dr., Ste. 510, Arlington, VA 22203, jfarrell@arctic.gov)

Scientific evidence indicates that global climate changes, and the most rapid changes are in the Arctic. Northern temperatures have increased twice as fast as elsewhere, the areal extent and thickness of Arctic sea ice have decreased dramatically, sea level rises as oceans warm and as glaciers and icesheets melt, and the ocean becomes more acidic as it absorbs atmospheric carbon dioxide. Changes in climate, as well as advances in technology, and the demand for resources have resulted in a newly accessible Arctic Ocean that invites to commercial fishing, shipping, tourism, mineral, and energy extraction interests and catches the attention of the US Coast Guard and the US Navy, among others. Many activities associated with these interests have implications for sound in the sea, and this presentation will present some of these at an overarching level. The United States has barely begun the baseline oceanographic research necessary to support national and international goals for ecosystem-based management and marine spatial planning in the Arctic Ocean. The US Arctic Research Commission, an independent federal agency, that advises Congress and the Executive Branch calls for Arctic Ocean research that would help policy makers address these issues.

9:25

1aAO2. The Arctic underwater acoustic environment. Christine Erbe (JASCO Appl. Sci., Brisbane Technol. Park, 1 Clunies Ross Ct, Eight Mile Plains, Queensland 4113, Australia, christine@jasco.com)

This talk gives an overview of the underwater acoustic environment in the Arctic. Sources of sound will be reviewed, including natural ambient noise (physical + biological) and man-made noise. Sound propagation characteristics will be discussed and examples from field studies will be presented. Features unique to the Arctic will be highlighted. Potential impacts of underwater noise on marine mammals will be summarized. Wild guesses will be made in an attempt to estimate how things might change in the future, particularly relating to climate change.

9:45

1aAO3. Soundscapes under sea ice: Can we listen for open water? Jeffrey Nystuen (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105 and HCMR, Attica, Greece, nystuen@apl.washington.edu) and Jennifer Miksis-Olds (Penn State Univ., State College, PA 16804)

Different geophysical conditions at the ocean surface generate distinctive sound spectra underwater, thereby allowing the physical environment to be monitored using passive acoustics. In the central Bering Sea, a passive aquatic listener (PAL) was deployed on the NOAA M5 mooring through the 2008 and 2009 winter seasons, allowing the ambient sound under sea ice to be recorded. Distinctive features of sea ice soundscapes include loud sound levels during periods of ice formation and very quiet high-frequency levels during periods of solid ice coverage. These soundscapes do have distinctive spectra features that distinguish them from other geophysical conditions, including wind-generated spectra, drizzle, and rainfall onto open water. This indicates that it is possible to listen for ice-free surface conditions. In turn, this will allow sub-surface ocean instrumentation on drifters, profilers, or moorings to detect open water and surface during relatively safe ice-free surface conditions (by listening), thus minimizing damage risk and allowing data transfer by satellite data links, an important advance for engineering oceanographic studies in ice covered seas. [Work sponsored by ONR Marine Mammals.]

10:05—10:30 Break

10:30

1aAO4. Acoustic monitoring of a melting Arctic Ocean. Peter N. Mikhalevsky (Acoust. and Marine Systems Operation, SAIC, 4001 N. Fairfax St., Arlington, VA 22203, mikhalevskyp@saic.com)

The waters of the Arctic Ocean have been warming since the early 1990s. Average maximum temperatures have risen by more than 1 °C. In the last 20 years submarine measurements of sea ice draft have shown a 40% reduction in average sea ice thickness while satellite remote sensing has shown a 14% reduction in sea-ice extent over the same period decreasing at a rate of 3%–5% per decade with thicker multi-year ice at 7%–10% per decade. Forecasts indicate that if these trends continue the Arctic Ocean could be ice-free in the summer before the end of this century. Significant effort is needed to expand our observational capabilities in the Arctic Ocean to support better modeling, forecasting, and improve our understanding of this critical ocean and the linkages to global climate. One technique acoustic thermometry has been shown to be a very effective for monitoring average heat content and average temperature in the Arctic Ocean and, in particular, in the Arctic Intermediate Water layer. Two experiments conducted in 1994 and 1999 measured the warming and demonstrated the feasibility of long term observations. Plans are in process to incorporate acoustic thermometry and tomography in *in-situ* Arctic Ocean observatories.

10:50

1aAO5. Improved assessment of Arctic and Antarctic zooplankton: From acoustic scattering models to a better understanding of polar ecosystems. Joseph D. Warren, Joy N. Smith (School of Marine and Atmospheric Sci., Stony Brook Univ., 239 Montauk Hwy., Southampton, NY 11968, joe.warren@stonybrook.edu), Patrick H. Ressler (Alaska Fisheries Sci. Ctr., NOAA, Seattle, WA 98115), and David A. Demer (Southwest Fisheries Sci. Ctr., NOAA, La Jolla, CA 92037)

Polar regions contain some of the most productive (and in some cases unspoiled) ecosystems on the planet. Zooplankton, particularly krill, are the preferred prey for numerous fish, bird, and mammal species including animals that are endangered, commercially fished, or ecologically important. Acoustic echosounders provide an excellent method for assessing the abundance and distribution of zooplankton over large areas (> 100 s km^2) through ship-based surveys and with high-temporal and spatial resolution. However, in order to accurately convert acoustic backscatter into a biologically meaningful measure (e.g., animal abundance, biomass, and calories), validated scattering models with well-understood input parameters are needed. This talk will present data from both Antarctic and Arctic (Bering Sea) zooplankton with regard to improvement of scattering models and the range and variability of model inputs for krill and other zooplankton, implications of these changes on acoustic estimates of zooplankton biomass or other metrics, and how these data may be used in ecosystem studies, particularly the relationships between zooplankton and their higher-trophic level predators.

Contributed Papers

11:10

1aAO6. Quantifying fish behavior with high-frequency scientific echo sounder and parametric sonar. Olav Rune Godø, Rolf J. Korneliussen (Inst. of Marine Res., P.O. Box 1870 Nordnes, N-5817 Bergen, Norway), and Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Schools of Norwegian spring-spawning herring (*Clupea harengus*) were observed in their wintering area off the northwest coast of Norway in November 2009 with the Simrad EK60 scientific echo sounder operating at 18, 38, 70, 120, and 200 kHz and Kongsberg TOPAS parametric sub-bottom profiling sonar, with nominal difference-frequency bandwidth 0.5–6 kHz. The beamwidths are 7 deg except at 18 kHz, where it is 11 deg. The TOPAS difference-frequency beamwidth decreases with both range and frequency, but lies within the range 3–5 deg at the depths of the detected schools, typically 50–300 m. The registered depths of the upper surfaces of the detected schools have been compared, indicating that the apparent depth often, but not always, is inversely related to beamwidth. The phenomenon seems to be related to the pattern of backscattering with respect to frequency within the same school, which appears to have two distinct modes. The results are analyzed and discussed relative to the phenomenon of gas release and avoidance reaction. [Work partly supported by Norwegian Research Council Grant No. 184705 and ONR Award No. N000140910482.]

11:25

1aAO7. Target strength spectrum of Atlantic herring over a low-kilohertz frequency band. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Olav Rune Godø, Rolf J. Korneliussen, and Eirik Tenningen (Inst. of Marine Res., Nordnes, N-5817 Bergen, Norway)

Atlantic herring (*Clupea harengus*) in the stock of Norwegian spring-spawning herring were observed in their wintering area off the northwest coast of Norway in November 2009. The volume backscattering coefficient s_V was measured with the Simrad EK60/38-kHz scientific echo sounder, yielding numerical values of fish density ρ through the fundamental equation of echo integration $s_V = \rho\sigma$, where ρ is the mean backscattering cross section of the ensonified fish. Derived values of ρ were used to convert measurements of s_V made with the Kongsberg TOPAS PS18 parametric sub-bottom profiling sonar to infer σ, thence its logarithmic measure target strength, over the approximate difference-frequency band 1.5–6 kHz. [Work supported by ONR Award No. N000140910482 and Norwegian Research Council Grant No. 184705.]

Session 1aBB

Biomedical Ultrasound/Bioresponse to Vibration: Diagnostic Applications of Ultrasound

Parag V. Chitnis, Cochair
Riverside Research Inst., 156 William St., New York, NY 10038-2609

Matthew W. Urban, Cochair
Mayo Clinic College Medicine, Dept. of Physiology and Biomedical Engineering, 200 First St., SW, Rochester, MN 55905

Chair's Introduction—7:55

Contributed Papers

8:00

1aBB1. Visualization of atherosclerotic plaque mechanical properties using model based intravascular ultrasound elastography. Michael S. Richards, Shayin Jing, and Marvin M. Doyley (Dept. Elect. and Comp. Eng., Univ. of Rochester, Rochester, NY 14627)

Cardiovascular disease related deaths occur primarily from fatty plaques inside an artery that rupture and cause clots that disrupt blood flow. The study of arterial plaque mechanics is essential to the monitoring of atherosclerosis and the detection of vulnerable plaques or the likelihood of their rupture. The aim of this work is to measure and image the mechanical behavior of plaques using a clinically available intravascular ultrasound system (IVUS). To test the methods, radial displacements were measured from simulated images and phantom data within two dimensional IVUS rf images using standard elastography techniques. Simulations and phantom experiments were designed to mimic arterial geometries and deformations, which include plaque regions of varying stiffnesses. In addition, an optimization inversion algorithm was used to infer the elastic modulus of the underlying material. This algorithm utilizes a finite element based model and assumes that the deformation is quasi-static, plane strain and that the material is incompressible and linear elastic. The full spatially reconstructed modulus images, with no geometric constraints, are presented and compared to radial strain images. The phantom results are compared to the known modulus contrasts, and the simulated images are compared to contrasts and distributions.

8:15

1aBB2. *In-vivo* time reversal blastography: A passive correlation tomography of complex shear wave field within in soft tissues. Stefan Catheline, Thomas Gallot, Philippe Roux, Javier Brum, and Carlos Negreira (Laboratoire de Gophysique Interne et Tectonophysique (LGIT), CNRS & Universit de Grenoble, France, stefan.catheline@obs.ujf-grenoble.fr)

Heartbeats, breathing, and muscle activity: the human body presents natural sources of vibration creating a complex shear wave field. Such a noise is demonstrated in this work to be useful for quantitative elasticity imaging. Indeed, the shear wave noise inside soft tissues is measured by ultrasonic techniques developed in elastography. The particle motions estimated along lines or planes inside the medium can be used as artificial stars for time reversal experiments. In another language, Green's functions can be recovered from noise correlations which results in a totally passive imaging technique. The use of shear wavelength or phase speed to reconstruct an elasticity mapping will be discussed.

8:30

1aBB3. Time reversal elastography: A correlation tomography of complex shear wave field in soft solids. Thomas Gallot, Stefan Catheline, Philippe Roux, Javier Brum, and Carlos Negreira (Laboratoire de Gophysique Interne et Tectonophysique (LGIT), CNRS & Universit de Grenoble, France, thomas.gallot@obs.ujf-grenoble.fr)

Shear waves in a soft solid can be focalized using time-reversal (TR) methods. The shear velocity, directly related to the elasticity of the medium, can be evaluated from the TR field. This study proposes to perform quantitative imaging of the mechanical properties inside a soft solid. Shear waves are created by finger impacts all over the surface of the solid. One component of the vectorial field inside the solid is measured by ultrasonic techniques developed in elastography. Inspired from Green's function retrieval methods, the TR field is computed from cross-correlation of this random-like displacement signal. One point is chosen as a virtual source where the spatio-temporal refocusing is observed. Thus the method allows focalization of shear wave everywhere inside the solid without any source embedded in the solid. The local shear velocity is then deduced from the TR field. Finally, a two dimensional velocity map of a bi-layer medium is performed.

8:45

1aBB4. Clutter and sources of image degradation in fundamental and harmonic ultrasound imaging. Gianmarco Pinton (Institut Langevin, Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, Paris 75005, France, gianmarco.pinton@gmail.com) and Jeremy Dahl (Duke Univ., Durham, NC)

Although harmonic imaging is used extensively, the mechanisms for image quality improvement are still poorly understood. Complex simulations or experiments that incorporate the effects of tissue heterogeneities and multiple scattering are required to describe image quality metrics such as clutter noise and less sensitivity to phase aberration. A finite difference simulation of ultrasonic propagation in a medium with heterogeneities in nonlinearity, attenuation, density, and speed of sound is presented. We investigate scattering from sub-wavelength targets and from macroscopic tissue structure. The simulated acoustic response from a single sub-wavelength scatterer is compared to analytical results for different scatterer sizes. This numerical implementation is then used to model tissue in the human body based on the visible human project and insonication from a from a 2.1-MHz diagnostic transducer. There is a close agreement between the analytical Rayleigh distribution, where the scattered intensity is proportional to f^4, and simulation. The primary source of image degradation in the fundamental PSF comes from reverberation in the near-field abdominal structures. Phase aberration is the largest source of clutter in the harmonic PSF. Clutter in the PSFs occurs primarily from low-level contributions distributed over a large area.

9:00

1aBB5. Feasibility study of superharmonic imaging using chirps. Paul L. M. J. van Neer, Mike G. Danilouchkine, Guillaume M. Matte (Dept. of Biomedical Eng., Erasmus Medical Ctr., Rm. Ee 23.32, Dr. Molewaterplein 50, 3015 GR Rotterdam, The Netherlands, p.vanneer@erasmusmc.nl), Martin D. Verweij (Delft Univ. of Technol., Mekelweg 4, 2628 CD Delft, The Netherlands), and Nico de Jong (Erasmus Medical Ctr., 3015 GR Rotterdam, The Netherlands)

Superharmonic imaging (SHI) targets a combination of the third to fifth harmonics. It was proven to have certain advantages in comparison with the established imaging standards in medical ultrasound. SHI enhances the spa-

tial resolution and improves the quality of echographic images, mainly by eliminating reverberation artifacts at the chest wall. However, SHI suffers from ripple artifacts, originating from the spectral gaps in between harmonics, and degrading the temporal resolution. To solve this a chirp-based SHI protocol was employed and its characteristics investigated, i.e., point spread function (PSF). The protocol was implemented for an interleaved phased array probe (44+44 elements tuned at 1.0+3.7 MHz), connected to a fully programmable ultrasound system. A linear chirp (center frequency 1 MHz; bandwidth 40%) was used for excitation. To obtain the PSF, the rf traces were recorded at focus along the lateral axis and convolved with the decoding signal. This was computed using KZK simulations. A PSF comparison between a superharmonic chirp and the third harmonic of a 2.5-cycle Gaussian apodized sinus burst at 1 MHz showed a decrease in axial pulse length of 46% at −6 dB and 32% at the −20 dB level in favor of SHI. Chirp based SHI is virtually free of ripple artifacts and therefore feasible.

9:15

1aBB6. Kramers–Kronig relationships applied to shear wave propagation in soft tissues. Matthew W. Urban, Shigao Chen, Carolina Amador, and James F. Greenleaf (Dept. of Physio. and Biomedical Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, urban.matthew@mayo.edu)

Multiple methods have been proposed to characterize the viscoelastic material properties of soft tissues to differentiate normal from pathological tissue. Shearwave dispersion ultrasound vibrometry is a method that uses the dispersion of the shear wave velocity to characterize tissue shear elasticity and viscosity. It has recently been shown [M. W. Urban and J. F. Greenleaf, Phys. Med. Biol. **54**, 5919–5933 (2009)] that the Kramers–Kronig relationships can be applied to shear wave measurements in *ex vivo* skeletal muscle over a frequency bandwidth of 100–600 Hz. We will describe this process where shear wave attenuation is modeled using a power law, and the shear wave velocity dispersion is predicted by applying the Kramers–Kronig relationships. In this study, we apply this method to shear wave propagation measurements made in different soft tissues such as *ex vivo* porcine skeletal muscle and kidney and *in vivo* porcine liver. The shear wave velocity predictions using the Kramers–Kronig relationships agree with the velocity measurements. The mean square error for these tissues ranged from 0.0024–0.0625 m^2/s^2. The Kramers–Kronig relationships can be used in shear wave propagation measurements in different types of soft tissues. [This work was supported in part by NIH Grants Nos. EB002167 and DK082408.]

9:30

1aBB7. Visualizing the functional properties of life-threatening atherosclerotic plaques using targeted ultrasound contrast agent and intravascular ultrasound. M. M. Doyley, H. Shekhar (Dept. of Elec. and Comput. Eng., Hopeman Eng. Bldg., Univ. of Rochester, Rochester, NY, doyley@ece.rochester.ed), J. S. Allen (Univ. of Hawaii-Manoa, Honolulu, HJ), and J. Rychak (Targeson, Inc., La Jolla, CA 92037)

Acute coronary syndromes may occur when life-threatening atherosclerotic plaques rupture in the later stages of advance cardiovascular disease. Coronary angiography is currently the gold standard for assessing the severity of coronary disease; however, its efficacy is questionable. We have developed a prototype intravascular ultrasound (IVUS) system to assess the functional properties (i.e., inflammation or angiogenesis or both) of atherosclerotic plaques and the arterial wall. In this paper, we report on the results of flow studies that were conducted to assess the nonlinear acoustic behavior of a novel perflourocarbon ultrasound contrast agent, Targestar, at high-transmit frequencies (30 and 40 MHz). Pulsatile flow studies were conducted with vessel phantoms. The subharmonic response of the targeted and non-targeted agents by applying an eight-order band-pass filter to the digitized radio-frequency echo frames. We compared the fundamental and subharmonic signals in absolute and relative terms using paired t-tests. SHI IVUS improves the visualization of the plaque relative to the main vessel, and the subharmonic behavior of the bubbles compared favorable to our theoretical expectations.

9:45

1aBB8. Estimating ultrasound attenuation along the propagation path using a reference phantom. Yassin Labyed and Timothy Bigelow (Dept. of ECpE, Iowa State Univ., 2215 Coover Hall, Ames, IO 50011)

The objective of this study is to develop an algorithm that uses backscattered data to estimate the total attenuation of ultrasound along the propagation path. The sample and a reference phantom are scanned using the same transducer and same power settings. The power spectrum of the gated window in the ROI of the sample is divided by the power spectrum of the reference phantom. The resulting ratio is then multiplied by four Gaussian filters at four different center frequencies to obtain four new spectra. The center frequency of each new spectrum is equal to the center frequency of the Gaussian filter shifted by a term that is proportional to the total attenuation. The total attenuation is calculated by monitoring the shift in the center frequencies of the new spectra with respect to the center frequencies of the Gaussian filters. The algorithm is tested using rf data generated by computer simulations. The sample and the reference had attenuation coefficients 0.7 and 0.5 dB/cm MHz, respectively. The error in the attenuation estimate is less than 5% for an ROI that is five pulse lengths long and contains 50 independent beam lines. [Work supported by NIH Grant CA111289.]

10:00—10:30 Break

10:30

1aBB9. Effective scatterer size estimation based on mathematical form factors in a rat fibroadenoma model. Lauren A. Wirtzfeld (Bioacoustics Res. Lab., Elec. and Comput. Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL, 61801, lwirtz@uiuc.edu), Zachary T. Hafez, Ellora Sen-Gupta, Andrew P. Battles, Rita J. Miller, Sandhya Sawarte, Michael L. Oelze (Univ. of Illinois, Urbana, IL, 61801), Timothy J. Hall (Univ. of Wisconsin, Madison, WI, 53705), and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL, 61801)

Quantitative ultrasound imaging is a model-based approach aimed at lesion detection and classification. In this study, the backscattered rf from rat fibroadenomas was fitted to two mathematical models to yield effective scatterer diameter (ESD). The goal was to aid in understanding potential sources of scattering in live rat tumors. Data were acquired using three single-element transducers with center frequencies of 3.5, 7.5, and 13 MHz. Reference scans were acquired from a Plexiglas plate. Two different theoretical models of scattering were fitted to the data; a fluid filled sphere model, which does not support shear waves, and the spherical Gaussian model for the scatterer. Twenty histologically confirmed fibroadenomas were analyzed. The ESD values obtained spanned the range of typical acini sizes observed in the histology of the fibroadenomas, from approximately 50 μm with the 13-MHz transducer up to 200–300 μm with the 3.5-MHz transducer. The estimates from the Gaussian model were consistently higher than for the fluid filled sphere. The estimates of ESD decreased with increasing frequency, and microstructures corresponding to these different scales were observed in histology. [Work supported by NIH Grant R01CA111289.]

10:45

1aBB10. Inter-comparison of ultrasonic backscatter coefficient measurements of live rat fibroadenomas across multiple imaging platforms. Lauren A. Wirtzfeld (Bioacoustics Res. Lab., Elec. and Comput. Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, lwirtz@uiuc.edu), Goutam Ghoshal, Zachary T. Hafez (Univ. of Illinois, Urbana, IL 61801), Kibo Nam (Univ. of Wisconsin, Madison, WI 53705), Yassin Labyed (Iowa State Univ., Ames, IA 50011), Janelle A. Anderson, M-Teresa Herd (Univ. of Wisconsin, Madison, WI 53705), Rita J. Miller, Sandhya Sawarte (Univ. of Illinois, Urbana, IL 61801), James A. Zagzebski (Univ. of Wisconsin, Madison, WI 53705), Timothy A. Bigelow (Iowa State Univ., Ames, IA 50011), Timothy J. Hall (Univ. of Wisconsin, Madison, WI 53705), Michael L. Oelze, and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

The backscatter coefficient (BSC) as a function of frequency is a system and operator independent parameter. It is the basis for some quantitative ultrasound (QUS) based on spectral analysis being translated into clinical use. This study aims to extend previous work in well-characterized physical phantoms to live animals where tissue properties are unknown. Six Sprague Dawley rats with spontaneous mammary tumors (five fibroadenomas and one carcinoma) were imaged with three clinical systems (Zonare Z.one, Ul-

trasonix RP, and Siemens S2000) and one single element laboratory system. Data were acquired from approximately the same region of the tumor with each scanner using independent setups. Scans of a reference phantom and Plexiglas plate were acquired for the clinical and laboratory systems, respectively. The data were analyzed using methods developed by the respective research group. BSC versus frequency plots show agreement in magnitude and trend among the different systems. The BSC estimates overlap each other, showing no more variability between systems than between consecutive imaging planes demonstrating agreement across multiple platforms and varied processing techniques between laboratories. Results support the potential to base diagnoses on QUS parameters. [Work was supported by NIH Grant R01CA111289.]

11:00

1aBB11. Array measurement and imaging of sound transmission through the lungs. Zoujun Dai, Ying Peng, and Thomas J. Royston (Univ. of Illinois at Chicago, Chicago, IL 60607, daizoujun@gmail.com)

Non-contacting scanning laser Doppler vibrometry (SLDV) was used to measure torso surface motion at an array of points caused by sound introduced into the lungs in human subjects. The SLDV serves as a gold standard as it does not have the complexity of surface contact dynamics. Additionally, a sensor array integrated into a compliant mesh fabric chair was constructed and tested as a less expensive and faster means of acquiring similar acoustic array data. Silicone gel-coupled, piezoelectric contact sensors for the detection of subsurface acoustic sources at the skin surface were used in the array. Array measurements were processed to determine the response at each measurement point as a function of frequency up to 1.6 kHz. Acoustic images were generated indicating the amplitude level at each point relative to the amplitude measured at the suprasternal notch, which served as the reference input. The acoustic measurements illustrated the complexity of sound transmission in the human lung and chest. Expected attenuation over the scapulae and below the diaphragm was observed. There were other significant variations as a function of location and frequency; measurement results are discussed. [Work supported by NIH Grant # EB008373.]

11:15

1aBB12. Simulations for investigating contrast mechanism of biological cells with high-frequency scanning acoustic microscopy. Yada Juntarapaso (Graduate Program in Acoust., The Penn State Univ., Univ. Park, PA 16802) and Richard L. Tutwiler (The Penn State Univ., State College, PA 16804-0030)

Scanning acoustic microscopy (SAM) is one of the most powerful techniques for nondestructive evaluation and it is a promising tool to characterize the elastic properties of biological tissues/cells. Exploring a single cell is important since there is a connection between single cell biomechanics and human cancer. SAM provides advantages for investigating cells in the fact that it is non-invasive, provides elastic properties information in sub-cellular details, and no chemical stain on the cells. The first goal of this research is to develop a program for simulating the images and contrast mechanism obtained by high-frequency SAM. We focus on the mechanical and acoustical properties of HeLa cells, which can be analyzed from the $V(z)$ response. Two new algorithms for simulating $V(z)$ responses involve calculation of reflectance function for coupling medium/cell/substrate system, and are constructed based on ray theory and wave theory (angular spectrum). The second goal is to design/optimize transducer arrays for SAM. One-dimensional and two-dimensional (2-D) linear array and phased arrays are simulated using FIELD II program to analyze image resolution and volumetric imaging capabilities. The new 2-D array design will enhance the performance of SAM by electronically scanning and by providing a four-dimensional image of the cells development phase.

MONDAY MORNING, 19 APRIL 2010 KOZMOS, 9:30 A.M. TO 12:15 P.M.

Session 1aMU

Musical Acoustics: String Instruments

Paul A. Wheeler, Chair
Utah State Univ., Electrical and Computer Engineering, 1595 N. 1600 E., Logan, UT 84341

Invited Papers

9:30

1aMU1. Traditional string instruments of the world. Paul A. Wheeler (ECE Dept., Utah State Univ., 4120 Old Main Hill, Logan, UT 84322-4120, paul.wheeler@usu.edu)

The string instruments best known to musicians, educated in western music, are the orchestral strings (violin, viola, cello, and double bass) with the addition of the guitar and harp. This represents a very small sample of string instruments used throughout the world. This paper is intended to introduce the variety of string instruments used in traditional music of world cultures. The instruments include the following: (1) musical bows—the berimbau of Brazil; (2) African and European lyres—the krar of Ethiopia and crwth of Wales; (3) harps—the ennanga of Uganda, saung gauk of Burma, and the Paraguyan harp; (4) plucked lutes—the charango of the Andes of Arabia, the balalaika of Russia, sitar of India, and the pipa of China; (5) fiddles (bowed lutes)—the hurdy-gurdy and nyckelharpa of Europe, the imzad of North Africa, the sarangi of India, and the morin khuur of Mongolia; (6) plucked zithers - the valiha of Madagascar, kantele of Finland, dan bau of Vietnam, and the zheng of China; and (7) dulcimers (hammered zithers)—the cimbalom of Hungary, santoor of Iran, and the yangqin of China.

10:00

1aMU2. Contemporary approaches to modeling violin tone, dynamics, and response. F. J. Denaro (Dept. of Biology, Morgan State Univ., 1700 East Cold Spring Ln, Baltimore, MD 21251)

Long-time-average spectral graphs of violin tone reveal some important characteristics of excellent instruments. Included are good dynamic range and frequency response. Significantly there is a marked response in the 194–649 Hz range (high sonority), a lower response for 649–1296 Hz (nasal), good response for 1830–4342 Hz (brilliance), and low response in the 4340–6300 Hz (harsh) [H.

Dunnwald, Catgut. Acoust. Soc. J. **1**, 1–5 (1991)]. In this talk a working protocol on how to modify the frequency response in each of these areas is explained. Important insights obtained from the burnish/hydrate (B/H) method will be used to illustrate how this can be done in a systematic fashion. The B/H method produces reversible stiffness changes in the plate area of choice. As a result it is possible to evaluate the change in tone and response by listening or by acoustical measurements. If the change is a desired one, it can be made permanent by removing the wood. While the frequency ranges noted above display a degree of interdependence, addressing them independently has led to a general pattern on where the needed plate and corpus modifications should be placed for an excellent violin.

10:30—10:45 Break

Contributed Papers

10:45

1aMU3. High-speed imaging techniques for harpsichord plucking action. Marshall J. Brown (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, mjbrown3@uno.edu), C.-Y. Jack Perng (Stanford Univ., Stanford, CA), Thomas D. Rossing (Stanford Univ., Stanford, CA), and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA)

Harpsichords contain vibrating strings whose motion has been studied extensively. The harpsichord sound is produced by a plectrum plucking the string, a behavior not well modeled at present. A high-speed video camera is used to image sequentially the plucking action of a harpsichord plectrum of modern plastic. This action is quite complicated and highly impulsive, too fast for the human eye to register. The tongue holding the plectrum moves backward in the escapement as well as when the plectrum slips by the string as the jack returns to its rest position. The plectrum may contact the string more than once during these events. The high-speed camera is a Casio Exilim Pro EX-F1, capable of 1200 frames/s. To study the plectrum action, monochords and harpsichords were used. One of the major difficulties with high-speed camera images provides sufficient light while maintaining CMOS saturation and keeping image coherence maximized for the fast shutter action. Various solutions to this problem were tested, including using direct sunlight. Further problems encountered (such as CMOS color response) will be described and their solutions discussed. Images show the plucking, the escapement, the plectrum motion, and the string excitation in detail during the plucking action.

11:00

1aMU4. Toward the modeling of harpsichord plucking. Chao-Yu J. Perng (Dept. of Phys., Stanford Univ., 382 Via Pueblo Mall, Stanford, CA 94305, perng@stanford.edu), Thomas D. Rossing (Stanford Univ., CA 94305), Marshall J. Brown, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The harpsichord's plucking action is much less studied than the piano's hammer striking action. A theoretical model was first proposed [D. H. Griffel, J. Sound Vib. **175**, 289–297 (1994)] treating the plectrum as a rigid body or two rigid bodies connected with a torsional spring. This was later investigated experimentally [N. Giordano and J. P. Winans II, J. Sound Vib. **224**, 455–473 (1999)] using a capacitive sensor, resulting in a more refined model based on simple one-dimensional beam dynamics. In our work, we employ various string position sensors, and a more detailed plucked string trajectory is captured and analyzed. Also aided by high-speed camera images, studies on the detailed plectrum's shape of deflection are done. By a more rigorous treatment of elastic beam theory and looking into interactions of the plectrum and string after the initial release, we present a revised model on the plucking mechanism of the harpsichord string that contributes to its characteristic tone quality.

11:15

1aMU5. Haptic interaction for music with unrestricted freedom of the musician's motion: Promoting augmented interaction with physical tools. Edgar Berdahl, Gunter Niemeyer, and Julius O. Smith, III (Dept. of Music, CCRMA, Stanford Univ., Stanford, CA 94305, eberdahl@ccrma.stanford.edu)

A haptic musical instrument is an electromechanical musical instrument that provides a musician not only with audio feedback but also with force feedback. Many commercial force feedback devices consist of a tool, such as a thimble or pen, attached to the end of a robotic arm. The musician can grab onto the tool and use it to manipulate a virtual musical instrument. While this configuration provides multiple degrees of freedom, it is suboptimal for many musical interactions. For instance, typically the workspace of the arm is restricted, which can detract from the musician's enjoyment of the experience. The workspace restriction can be resolved by freeing the tool from the end of the haptic device and having the musician hold the tool, which may be a plectrum, bow, or drumstick, in his or her hand. Further benefits can be reaped if multiple points of haptic interaction are available, enabling the musician to manipulate the points and move freely in between them. A dual voice coil woofer provides a convenient collocated sensor /motor for constructing a single degree-of-freedom point of haptic interaction. We present an example instrument in which a musician employs a conventional plectrum to interact with a virtual digital waveguide string.

11:30

1aMU6. Acoustic changes produced by hydration and burnishing of violin plates. F. J. Denaro (Dept. of Biology, Morgan State Univ., 1700 East Cold Spring Ln, Baltimore, MD 21251) and M. R. Serbyn (Morgan State Univ., Baltimore, MD 21251)

To achieve the proper balance of tone and loudness throughout the frequency range of a fine violin, various methods of plate tuning have been proposed. Unfortunately, their effectiveness can be judged only *a posteriori*, for even if a violin plate response is well characterized, where to work the wood is poorly known. What is needed is a process that does not remove wood until the desired acoustical characteristics are produced regardless of the method used. The technique proposed by us consists of systematic hydrations and/or compressional burnishing of focal areas of the violin plates or corpus. Hydration is acoustically equivalent to a decrease in stiffness, while burnishing increases stiffness. Both are reversible without marked changes to the wood. In order to establish how these treatments affect the material constants of the plates, strips of maple and pine were tested in both treated and untreated states and the changes noted. Results suggest temporary changes to plate stiffness. The magnitude of these changes is sufficient to affect acoustic signatures and holographic patterns and hence can be detected. The accumulative effects of such changes to violin tone are noted in our companion paper.

11:45

1aMU7. Systematic optimization of violin tone by an acoustically reversible approach: The burnish/hydration technique. F. J. Denaro (Dept. of Biology, Morgan State Univ., 1700 East Cold Spring Ln, Baltimore, MD 21251) and M. R. Serbyn (Morgan State Univ., Baltimore, MD 21251)

The companion paper supports the hypothesis that focal alterations to the stiffness of the wood can be produced by compressional burnishing or hydration of discreet areas of the plates or corpus. The effects of both burnishing and hydration are reversible. Therefore, one has real-time-feedback for modeling optimal violin tone. As a result only those acoustic changes which are desired are made permanent by removal of wood. To develop proof of concept we applied this approach to both top and back plates and to violins in the white. The experiments consisted of evaluation of tap tone frequencies and evaluating violins played in the white. Standard recording- and signal-analysis equipment was used. The conclusions drawn from measurements correlated well with those from listening, that is, incremental changes in the tone produced by this technique can guide the carving of the plates. An important strength of the technique is that it can be applied with other

methods of plate tuning, namely, tap-tone methods, holographic techniques, computer analysis, etc. The approach should provide a valuable adjunct in making instruments of the violin family.

12:00

1aMU8. Estimation of loop filter coefficient for the silk stringed instruments. Sangjin Cho and Uipil Chong (School of Elec. Eng., Univ. of Ulsan, Ulsan, Republic of Korea, sjcho75@ulsan.ac.kr)

It is very important to represent an exact attenuation of a sound in a physical modeling of the plucked string instruments. Vlimki *et al.* proposed a simple one-pole low-pass filter called loop filter to depict the attenuation of the sound in the extended KS algorithm. In order to estimate the filter coefficient, attenuations and gains of each harmonic are measured and calculated, respectively. The coefficient can be obtained by using the mean squared error between magnitude response of the filter and the gains of each harmonic. In this case, unlike steel stringed instruments, the accuracy of filter coefficient estimation for the silk stringed instruments depends entirely upon choosing the specific region of the recorded sound. In this paper, 84 different notes of the silk stringed instrument called Gayageum, Korean traditional plucked-string instrument, were recorded, and estimation of filter coefficient was performed for the following regions: entire, decay, sustain, and RT1 (reverberation time) region. The estimation results were good in order of decay, RT1, sustain, and entire region. Especially, the performance for the decay region was very good. [This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. R01-2008-000-20493-0).]

MONDAY MORNING, 19 APRIL 2010 GRAND BALLROOM V, 8:00 TO 9:00 A.M.

Session 1aNCa

NOISE-CON: Plenary

Patricia Davies, Chair
Purdue Univ., School Mechanical Engineering, 140 Martin Jischke Dr., West Lafayette, IN 47907-2031

Chair's Introduction—8:00

Invited Paper

8:05

1aNCa1. Sonic boom: From bang to puff. Kenneth J. Plotkin (Wyle Labs., 241 18th St. S., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wyle.com)

For almost 40 years commercial supersonic flight over land has been prohibited because of the negative impact of sonic booms. During that time, technology advances have raised the possibility of a muted sonic boom—a sonic "puff"—that could be acceptable over populated areas. This paper reviews the physics of sonic boom generation and propagation, and why booms from conventional aircraft are usually *N*-waves with unacceptably large shock waves. The low-boom shaping technology that can minimize shocks, thus reducing sonic boom loudness to acceptable levels, is presented. The success of the DARPA/NASA Shaped Sonic Boom Demonstrator in validating shaping theory is discussed. Concepts for attaining practical aircraft configurations, versus ideal optimal shapes, are described.

8:45—8:50 Questions

8:50—9:00 Announcements

Session 1aNCb

NOISE-CON: Materials for Noise Control—Manufacturer Presentations

Stephen I. Roth, Chair
Roth Acoustical Associates, 2352 Norton Rd., Pittsburgh, PA 15241

Contributed Papers

9:15

1aNCb1. Effect of intrinsic parameters on sound absorption and transmission loss: A parametric study. Paresh Shravage, Sachin Jain, and Nagesh Karanth (Automotive Res. Assoc. of India, Noise Vib. and Harshness Lab, Automotive Res. Assoc. of India (ARAI), S. No 102, Vetal Tekadi, Off Paud Rd., Kothrud, Pune 411038, India pareshshravage@gmail.com)

Sound absorbing materials are indispensable nowadays for noise control treatments in transport industry. The acoustical behavior of these acoustic materials is governed by five physical (e.g., porosity, flow resistivity, tortuosity, viscous characteristic length, and thermal characteristic length) as well as three mechanical parameters (e.g., Young's modulus, Poisson ratio, and loss factor). The characterization of these porous materials is very crucial as it plays an important role in design and development stage itself for predicting acoustic behavior of multilayer porous materials for higher sound absorption and transmission loss. This prediction depends on measurement accuracy of macroscopic physical parameters which are very difficult to measure except porosity and flow resistivity, which is the only standardized test until today; also availability of such rigs is also a problem for manufactures as they are available only at specialized test laboratories. This paper presents effect of all of these parameters on sound absorption coefficient and transmission loss of the porous materials using Johnson—Champoux—Allard and Biot model for poroelastic materials. It also discusses results of simulation with effect of each parameter on acoustic behavior of the sound absorbing materials.

9:30

1aNCb2. A comparison of analytical and optimization inverse techniques for characterizing intrinsic parameters of porous materials. Paresh Shravage, Sachin Jain, and Nagesh Karanth (Noise Vib. and Harshness Lab, The Automotive Res. Assoc. of India (ARAI), Survey No. 102, Vetal Hill, Off Paud Rd., Kothrud, Pune 411 038, India, pareshshravage @gmail.com)

In recent years, sound absorbing materials are finding many applications in transport industry. In automobiles, they are used as noise control treatments in engine compartment as well as in passenger cabin. To understand and simulate acoustic behavior of these noise control treatments, intrinsic physical parameters are required. The acoustical behavior of poroelastic material is governed by five macroscopic intrinsic parameters, e.g., porosity, flow resistivity, tortuosity, and characteristics lengths as well as three mechanical parameters. Out of these five physical parameters, porosity and flow resistivity can be measured directly by available standardized methods. But measurement of physical parameters such as tortuosity, viscous, and thermal characteristic lengths is very difficult and no accepted procedure is available for their measurement. As an alternative, analytical inverse approach (mid-frequency) and optimization technique Genetic algorithm are well known in the literature. This paper compares the results from both techniques for intrinsic parameters. It also presents the effect of inverted parameters on sound absorption and transmission loss of the porous materials using Johnson–Champoux–Allard model.

9:45

1aNCb3. Hand troweled and spray applied sound absorbing acoustical plasters. Howard Podolsky and Andrew Sarcinella (Pyrok Inc., 121 Sunset Rd., Mamaroneck, New York 10543, howard@pyrok.com)

This paper will discuss the types of hand applied and spray applied sound absorbing acoustical plasters manufactured by Pyrok Inc. Pyrok Acoustement products are spray applied acoustical plasters which are formulated with gypsum plaster or Portland cement. The Portland cement Acoustement 40 can be used for exterior as well as interior applications. Acoustment Plaster 20 and Acoustment Plaster 40 are gypsum based and are for interior applications only. The Pyrok StarSilent System has a smooth hand trowel applied finish that looks like monolithic drywall as well as a superfine sprayed finish. This paper will also discuss ASTM C423 testing and how it relates to spray and hand applied acoustical plasters. Although all spray applied and hand applied sound absorbing products are tested to ASTM C423, there are differences and inconsistancies when evaluating and comparing their sound absorption data.

10:00

1aNCb4. Numerical study of noise transmission through functionally graded panels. Changzheng Huang and Steve Nutt (Dept. of Chemical Eng. and Mater. Sci., Univ. of Southern California, 3651 Watt Way, VHE 602 Los Angeles, CA 90089-0241, changzh@usc.edu)

Functionally graded panels are panels made of either functionally graded materials (FGMs) or functionally graded structures. FGM panels are traditionally used as thermal insulation barriers and they find applications in space and launch vehicles. Functionally graded structures include foams with continuous variations in pore size from one side to the other side. Such foams are incorporated as the core layer in sandwich panels. The latter is often used as partition walls in the building and transportation industries. In this study, we are concerned with the noise transmission characteristics of functionally graded panels. We developed an analytical model to derive the transfer matrix for the functionally graded medium. Using this model, we were able to make numerical predictions for sound transmission through functionally graded panels. The gradation parameter provides an additional degree of freedom in controlling the noise transmission characteristics. This is helpful in maximizing the noise reduction in panel design. Furthermore, our transfer matrix model applies to conventional discretely layered panels as well. In fact, the latter is just a special case in our modeling formulation. Numerical examples and results are presented and discussed.

10:15—10:30 Break

10:30

1aNCb5. Achieving minimum impact insulation class 50 rating using resilient clip technology in lightweight construction. Wilson Byrick (Pliteq Inc., 1370 Don Mills Rd., Unit 300 Toronto, ON M3B 3N7, Canada, wbyrick@pliteq.com)

Resilient clip technology is an alternative to resilient channel for supporting gypsum wallboard in fire rated wall and floor-ceiling assemblies. Lightweight wood frame construction presents challenges in meeting building code requirements for impact insulation class (IIC) and sound transmission class (STC) ratings in floor-ceiling assemblies. Through laboratory testing we were able to compare different floor-ceiling assemblies commonly used in multi-family construction. Two different structures were tested, an 18 in. open web truss, 24 in. o.c., and a 12 in. engineered joist (TJI) 24 in. o.c. Various finish floor coverings were used including ceramic tile, vinyl, and engineered wood. Structures achieved minimum IIC and STC 50 with

and without 3/4 in. Gypsum concrete. Resilient rubber underlayment of varying thickness can further improve the IIC and STC ratings of the assembly as can adding a second layer of gypsum wallboard.

10:45

1aNCb6. Quieting small centrifugal fans: A material manufacturer's view. Steven M. Brown and Dan LaForgia (The Soundcoat Co. Inc., 1 Burt Dr., Deer Park, NY 11729, sbrown@soundcoat.com)

Noise control treatments have been developed for small centrifugal fans in fan trays. The fan trays are employed in cooling mobile electronics equipment. Similar treatments can be employed in stationary equipment racks used in the information technology and telecom industries. The noise control treatments incorporate especially sculpted sound absorbing material. This paper discusses the considerations used in developing the treatments, the resulting sculpted design of the treatments, and the acoustical material used in producing the treatments. The achieved noise reductions are presented. Comments are made on the resultant improvements in airflow and accompanying cooling capacity.

11:00

1aNCb7. Heating, ventilation, and air conditioning flexible duct acoustical solutions. Neil Silverman (Flexmaster USA, Inc., 4545 Pine Timbers # 324, Houston, TX 77041, neil@flexmasterusa.com)

Flexmaster USA has been manufacturing the highest quality flexible duct available for commercial applications for over 30 years and has a wide array of selections. There are three very distinct choices, type 6, type 8, and our Medical Grade Triple Lock Acoustical that have been tested by Energistics Laboratory, which is a research and test facility dedicated to the HVAC industry. The testing for Flexmaster USA's type 6, SPUN BOND Nylon, type 8, chlorinated polyethylene (CPE) fabric, and model MGA TLM (aluminum corrugated flexible duct) has shown significant reduction in sound. In other words, they absorb unwanted sound, including the most important mid-range frequencies. Our products listed here will significantly reduce unwanted noise in all of your rooms. These products allows sound—not air flow—to permeate through the liner and become absorbed inside the insulation. Each 6 in. test specimen was tested both straight and with a 90 deg bend for velocities of 0 FPM, 500 FPM, and 1000 FPM and compared to hard metal duct at the same velocities. Diameters of 6 in. thru 16 in. were provided by Flexmaster USA directly to Energistics Laboratory for the testing that was concluded in July of 2006. Unique to others in the industry, Flexmaster USA manufactures flexible duct without the use of glues or adhesives. The mechanical-locked inner core maintains its shape

11:15

1aNCb8. Low-mass–high-transmission loss portable speech isolation room for very important person protection. Mark Rubino (Industrial Noise Control, Inc., 401 Airport Rd. North Aurora, IL 60542, MRubino @industrialnoisecontrol.com)

This case study examines the design, construction, and testing of a lightweight, easily transportable speech isolation room capable of providing substantial anti-eavesdropping countermeasures for use by high-level government and intelligence agencies as a secure meeting and communications center that can be readily deployed worldwide. The performance criteria requiring a minimum field STC-35 utilizing a modular, scalable mechanical design capable of repeated assembly without the use of tools, a maximum component weight of 2.65 lb/ft^2 a 2 in. maximum wall thickness, and the incorporation of active and passive ventilation required the development of a robust, multi-layered composite panel and joining system as well as ancillary components. Using a combination of non-metallic limp mass and rigid metallic barriers separated by decouplers of varying properties in an aluminum shell brake-formed to distort natural resonance, we were able to achieve significant transmission loss (up to 45 dB) in the speech range of frequencies equaling the performance of commercially available products of nearly twice the weight and bulk. Interior room acoustical characteristics were optimized by using modular demountable sound absorptive units.

11:30

1aNCb9. Modular room design. Craig D'Anna (Industrial Acoust. Co., 1160 Commerce Ave., Bronx, NY 10462, cdanna@industrialacoustics.com)

Modular acoustical panels have been used to construct noise control enclosures at least since the early 1950s. This paper presents examples of the early days of modular acoustical enclosures and illustrates the current state of the art where pre-engineered acoustical panels are used to construct acoustical environments for Music Practice, Broadcast Facilities, Recording Rooms, and Sensitive Compartmented Information Facilities in a shorter period of time and at a lower cost than traditional drywall construction. Numerous examples are presented illustrating how, through the use of pre-engineered acoustical panel systems, manufacturers are able to guarantee the acoustical characteristics of a space while maintaining the ability to expand, demount, and relocate the rooms at a later point in time. The benefits of a no-hassle installation and acoustic guarantee, as well as multiple options for fit and finish that will equal the best craftsman's fine handiwork, are presented.

11:45

1aNCb10. Pearl Palms concert theater. Steven Udolph (Tectum Inc., P.O. Box 3002, Newark, OH 43058, sudolph@tectum.com) and Tera Aurther (Badertscher Communls)

The Pearl Theater was designed and built to meet the demands of both the sophisticated local market and the demanding high-end tourist market in Las Vegas, NV. These markets will quickly dismiss any venue that does not deliver a superior performance, as will the performers that a venue like this must attract in the highly competitive environment, that is, Las Vegas. The Pearl delivers for the audience, the performers, and the owners. Studio X at the Palms is the largest and best-equipped recording facility ever built in Las Vegas. The Pearl Concert Theater is hard wired to the studio to allow artists to conveniently and cost effectively create albums of their live performances in the theater. For the owners to realize a return on their investment in these facilities, the acoustics in the 3294 sq.ft theater had to be impeccable. But the need for a sophisticated sound environment was just the beginning. The aesthetics of the 2400-seat theater were also a major priority. To address both concerns, Dean Roofing & Insulation installed 25000 sq. ft of Tectum 1 in. panels in all of the traffic ways and throughout the theater. The Tectum panels and the seats are the only sound absorbing materials used in this project.

Session 1aNCc

NOISE-CON, Noise, and Signal Processing in Acoustics: Active Noise Control

Marty Johnson, Chair
Virginia Tech, 143 Durham Hall, Blacksburg, VA 24061

Contributed Papers

9:15

1aNCc1. A novel delayless frequency-domain filtered-x least mean squares algorithm for active noise control. Jie Duan and Teik Lim (Univ. of Cincinnati, 730 Riddle Rd., Apt. 205N, Cincinnati, OH 45220, duanjieee02@gmail.com)

The delayless frequency-domain implementation of least mean squares (LMS) algorithm is attractive because it can eliminate undesired block delay in the signal path while still remain the advantages of the frequency-domain LMS algorithm. However, the conventional delayless frequency-domain LMS experiences difficulties in slow convergence when it is applied to harmonic signal control at some frequencies. In this study, an overlap-save implementation of the delayless frequency-domain LMS algorithm is proposed to overcome the convergence problem mentioned above. Simulations show that the proposed implementation has better performance than the original one either in the pure tone application or in the vehicle powertrain noise application.

9:30

1aNCc2. Active noise control of magnetic resonance imaging scanner using inverse modeling technique. Mingfeng Li, Brent Rudd, and Jing-Huei Lee (Mech. Eng., Univ. of Cincinnati, 584-D Rhodes Hall, P.O. Box 210072, Cincinnati, OH 45221-0072, limf@ucmail.uc.edu)

In active noise control applications, internal model control (IMC) is a strategy to convert from feedback to feedforward control. In this configuration, the optimal solution of the controller will be the inverse model of the secondary path. However, the inverse model of the secondary path may not always exist. In fact, for the non-minimal phase system, the inverse model is unstable. In this case, the non-causal or delayed version of the inverse model will be employed. In general, this approach is not suitable for active noise control if the objective is to treat broadband noise. However, for harmonic signal control, this is not an issue. Magnetic resonance imaging (MRI), a procedure for medical diagnosis and biomedical research, generates undesirable levels of acoustic noise emission. The noise is normally dominated by clusters of harmonics throughout the audio spectrum. Due to its tonal nature, MRI noise control is a candidate for IMC. In this paper, an active noise control scheme using the inverse modeling technique is proposed and studied for pre-recorded MRI noises obtained during operation of a 4T whole-body MRI scanner. The recorded MRI noise is played back through a stereo system in a sound quality room and controlled using a headset equipped with the proposed active controller. The results demonstrate almost 20 dB reduction can be achieved.

9:45

1aNCc3. Active noise control of multiple cooling fans. Ryan Rust, Jonathan Blotter, Scott Sommerfeldt, and Kent Gee (Brigham Young Univ., 435 Crabtree Bldg., Provo, UT 84602, ryrust@gmail.com)

Fans are used to dissipate heat from electronic devices to keep the components working properly but create noise as a byproduct. Multiple fans are often needed in computer servers and other types of electronic equipment to adequately cool them. The fan array increases the cooling rate but also increases the noise produced. In addition to the added noise, differences between fans due to manufacturing cause the fans to rotate at different speeds, creating an annoying beat frequency. Feed forward active noise control was studied for an exhaust mounted two fan array using near field error sensors and control speakers. The tonal noise content of the blade passage frequency and the first harmonic was controlled at the error sensors for each fan. The tonal noise content was also reduced in the far field and the beat frequency reduced.

10:00

1aNCc4. A physically motivated room reverberation enhancement system that is stable in any (passive) room. Edgar Berdahl, Gunter Niemeyer, and Julius O. Smith III (Dept. of Music, CCRMA, Stanford Univ., Stanford, CA 94305, eberdahl@ccrma.stanford.edu)

Prior active room reverberation enhancement systems have typically employed microphones, artificial reverberator filters, and loudspeakers to change the reverberant properties of a room. However, acoustic feedback from the loudspeakers to the microphones has had the potential to drive such systems unstable. In contrast, techniques from feedback control can be used to design a stable room reverberation enhancement system from the ground up. A feedback controller implements a passive connection between virtual acoustic spaces, which are realized using digital waveguide networks, and special transducers, which operate concurrently as microphones and loudspeakers. Because the feedback controller models a passive system, it is theoretically stable for arbitrarily large loop gains in any (passive) acoustic environment. As a consequence, the system does not suffer from "ringing tones" at high-loop gains in the same way as prior systems have suffered. Furthermore, the system does not need to be re-calibrated if the properties of the room change or even if moved to a whole new room. Finally, this method for designing room reverberation enhancement systems may generally result in more realistic reverberant sound because it implements the acoustical features of a system that could exist naturally in the physical world.

10:15

1aNCc5. An electroacoustic sound transmission system that is stable in any (passive) acoustic environment. Edgar Berdahl (Berdahl Innovations, 103 Cortland Ave., San Francisco, CA 94110), Dan Harris (Sennheiser Res. Lab., 3239 El Camino Real, Palo Alto, CA 94306), Gunter Niemeyer, and Julius Smith, III (Dept. of Music, CCRMA, Stanford Univ., Stanford, CA 94305)

Active electroacoustic systems are commonly employed to transmit sound from one location to another. For example, consider the following configuration: a person talks into a microphone, which produces an electrically amplified signal for driving a loudspeaker, and the loudspeaker induces an output acoustic signal allowing someone else at a more distant location to hear the person talking. Unfortunately, such system designs typically do not consider acoustic feedback, which can destabilize the system and result in "howling." In contrast, a feedback control system can transmit sound from one location to another using collocated microphone/loudspeaker transducers without the risk of howling. We design feedback controllers that model positive real (i.e., passive) systems, such as a spring and a gyrator. The spring controller essentially binds the diaphragms of the two transducers together. We relate the spring controller to the string connecting two tin cans together in the classical tin can telephone. Measurements are performed on a real feedback control system with two transducers. Because the feedback controller models a passive system, it is stable in any (passive) acoustic environment. [Sennheiser Electronic Corporation helped support this work.]

10:30

1aNCc6. Three-dimensional audio interface: A user's survey. Kim Abouchacra, Solara Sinno (Dept. of Otolaryngol.-Head and Neck Surgery, AUBMC, American Univ. of Beirut, 11-0236 Beirut, Lebanon, ks05@aub.edu.lb), and Tomasz Letowski (U.S. Army Res. Lab., Aberdeen Proving Ground, MD)

Increasing complexity of military vehicles, high noise levels, and the command and control demands of modern warfare put high physical and mental load on crews operating the vehicles. The need for indirect driving and simultaneous control of various robotic systems demands multisensory interfaces between the crew and the operated systems. One of the promising types of new interfaces is a three-dimensional (3-D) audio interface that presents warning signals and tactical messages at spatial locations associated with the content of the emitted signals or messages. However, not all the information can be displayed through a 3-D interface with needed resolution and sound quality and either too little or too much information can be equally detrimental to the user. This paper is the summary of data obtained through a user's survey administered to 105 tankers regarding expected functionality of 3-D interfaces in armored vehicles. The results of the survey strongly support the value of the 3-D audio interface for most of the vehicle's systems and for most operational conditions. The general results and the specific details of the survey will be discussed.

MONDAY MORNING, 19 APRIL 2010 GRAND BALLROOM VII/VIII, 9:15 A.M. TO 12:00 NOON

Session 1aNCd

NOISE-CON, Noise, and Architectural Acoustics: Building Design and Construction for Effective Acoustic Performance

Kenric Van Wyk, Chair
Acoustics By Design, 123 East Fulton, Grand Rapids, MI 49503

Contributed Papers

9:15

1aNCd1. Acoustical considerations in future building design processes. Tyrone Hunter (Balance Acoust., 7710 Hazard Ctr., Dr. E-102, San Diego, CA 92108, huntert3@asme.org)

"Sustainability" is the major buzzword in the building design and construction arena today. The context often talks about products that will result in a smaller initial and ongoing carbon footprint. This paper will look at the idea of sustainability as a process, covering the end-to-end design and construction process as well as the process of generating and managing building data during its life cycle. The advent of lean construction and building information modeling is an opportunity to make acoustics more inherent to the conceptual design and prevent it from being a casualty of the "value engineering". Significant questions do arise, however, with regard to the acoustical information needed for such processes. These questions include, but are not limited to, what information is needed, who will need it, and how will it be exchanged. The answers to these questions differ when considering the above-mentioned processes and efforts to be holistically efficient and sustainable. The answers to these questions will be integrated into how building design is taught with regard to acoustics other specialties.

9:30

1aNCd2. Classroom noise at the Applied Research Laboratory, Penn State. Marshall Long (Marshall Long Acoust., 13636 Riverside Dr., Sherman Oaks, CA 91423, mlacoustics@sbcglobal.net)

Session 3.8: Building Design and Construction for Effective Acoustic Performance Classroom Noise at ARL Penn State by Marshall Long The Applied Research Laboratory at Penn State is a well known center of acoustics education in the United States. When Marshall Long Acoustics was selected to be the acoustical engineer on the new ARL building, we were particularly careful to check every noise source. When we went out on the final inspection, the question arose, "Why is the background noise level NC 42 in a classroom when it should be under 30?" The answer was not so elementary my dear Watson.

9:45

1aNCd3. Acoustics of technology enabled collaborative learning environments. Gregory A. Coudriet and Jeffery E. Babich (The Sextant Group, Inc., 730 River Ave., Ste. 600, Pittsburgh, PA 15212, gcoudriet @thesextantgroup.com)

Collaborative classrooms and workstations are increasingly common in higher education buildings. These spaces present unique acoustical challenges that force designers to rethink traditional concepts and solutions. In student-centered classrooms, the lecturer is removed as the visual and acoustical center of attention. Instead, students are seated in groups around technology enabled workstations, often facing away from the lecturer and other students. Acoustical conditions must support small-group interaction, classroom discussion, and learning via multimedia content. Outside the classroom, collaborative workstations vary greatly, but typically facilitate small-group interaction around a multimedia display and/or work surface. To promote impromptu use by students, these spaces are often open to highly public areas of learning commons, classroom buildings, and digital libraries. Traditional noise control solutions fail to isolate these spaces without compromising their spontaneous feel and function. Despite obvious acoustical challenges, both space types will continue to proliferate and new acoustical solutions must be developed.

10:00

1aNCd4. Small deviations and big failures in vibration and noise isolation. Byron Davis and Ahmad Bayat (Vibro-Acoust. Consultants, 490 Post St., Ste 1427, San Francisco, CA 94102, byron@va-consult.com)

The difference between design intent and actual implementation is frequently a problem in real constructions. For vibration and noise isolation hardware, seemingly innocuous deviations can result in major deficiencies in performance. From hardware selection to installation and adjustment, vibration isolation systems present significant challenges. In this paper, we offer insight into a broad spectrum of problematic rotating mechanical systems. Numerous categories of vibration-isolation failure are identified, and detailed photographs highlight subtle errors in isolator selection, installation, and adjustment that can lead to poor vibration and noise performance. Before-and-after vibration data are presented to illustrate the effect of poorly isolated rotating machinery. A spectacular example of comprehensive facility-wide failures is given, with ground vibration data from a brownfield site adjacent to an outdated textile plant.

10:15

1aNCd5. Retrofit structural insulated panels increase sound transmission loss of existing single family houses impacted by highway noise. Scott Harvey, PE (Phoenix Noise & Vib., LLC, 5216 Chairmans Ct., Ste. 107, Frederick, MD 21703-2881, sharvey@phoenixnv.com)

Owners of newly constructed houses complained of interior noise from a nearby interstate highway. The houses were constructed using structural insulated panels (SIPs) as opposed to conventional 2x4 wood studs. The SIPs are composed of 3.5 in. of expanded polystyrene insulation sandwiched between two layers of oriented strand board (OSB) and exhibit higher thermal efficiency than conventional construction. Sound transmission loss testing of the standard SIPs, however, reveals relatively low acoustical performance especially in the 630 Hz range. This characteristic acted as a band pass filter allowing only a portion of the broadband traffic noise to enter the house. The observed traffic noise inside the house had a clear tonal quality adding to the annoyance of the traffic noise impact. After consideration of several remedial options the builder decided to retrofit the exterior of the impacted houses using a combination of standard resilient channel, insulation, and additional OSB. Some window replacements were also required. These treatments resulted in reduced interior noise levels especially in the 630 Hz range.

10:30—10:45 Break

10:45

1aNCd6. The influence of openings in a partition on its of tonal sound transmission loss and some applications. Giora Rosenhouse (Technion-Israel Inst. of Technol., Prof. Giora Rosenhouse 89 Hagalil Str., Haifa 32684 Israel, fwamtech@bezeqint.net)

Kirchhoff's integral developed for optics [Born and Wolf 1970] and its approximations by Fraunhofer and Fresnel constitute a useful tool in acoustics, in general, as well as in building acoustics. They help in estimation of the changes in the sound level at the control point due to arbitrarily shaped solid barriers and openings in such barriers or partitions. While noise analysis justifies energetic approach, tonal sound is sensitive to frequency, phase, and interference effects. This dependence yields the mapping of positive and negative Fresnel zones over the barrier or the aperture. Those zones depend also on the location of the source and the control point. Since sound waves under regular conditions are linear and we assume locally reacting surfaces, then a small barrier or an aperture, which is located over a single Fresnel zone, causes a stronger acoustic effect at the control point than exactly the same elements if they cover parts of Fresnel zones of different signs. This situation allows, according to the need, acoustic design for minimization or maximization of the effect of apertures and barriers on the sound level at the control point. Applications involve arbitrarily shaped sound barriers, arbitrarily shaped openings in partitions and barriers, acoustic tiles, and transmission loss of non-uniform partition.

11:00

1aNCd7. Sound isolation provided by shading screens applied in facades. Jorge Matos and Antonio Carvalho (Lab. of Acoust., College of Eng., Univ. of Porto, 4200-465 Porto, Portugal, carvalho@fe.up.pt)

In most European countries, legislation exists about airborne sound insulation in dwellings, including facades. Mainly in southern European countries, the glazed windows of building facades normally have shading systems to minimize the excessive heating of interior rooms due to the solar rays' incidence and to provide for darkening of the room. The effect regarding sound isolation of those shading systems is usually not analyzed in the buildings' acoustic project. This study presents laboratory measured values of weighted sound isolation index (Rw) provided by several shading system types (outside and interior screens of different materials) and presents a simple model to predict their sound isolation index.

11:15

1aNCd8. Case study: Computer modeling of exterior mechanical noise from an existing building to develop mitigation for code compliance. Roman Wowk and Chris Papadimos (818 Fifth Ave., Ste. 207, San Rafael, CA 94901, roman@papadimosgroup.com)

This paper presents a case study of how computer modeling was used to develop exterior noise treatments for an existing building. Sound pressure measurements were used to first establish existing ambient noise levels and characterize building noise sources. Following that, a computer model of the building and surrounding area was constructed using commercially available environmental noise modeling software. Noise sources were modeled using sound power spectra from equipment manufacturers or approximated from measured sound pressure levels. The predicted ambient noise levels compared favorably to the measured data, thus validating the accuracy of the model. Once this baseline condition was established, a new building proposed between the existing building and the residences was added in the model to predict how it would function as a barrier to reduced noise at the residences. To develop additional noise treatments necessary for code compliance, the model was then used to determine the contribution of each piece of equipment at key receiver locations. Specific equipment could be targeted when considering noise treatments and noise exposure maps could quickly be produced for a combination of scenarios to aid in finding the most cost-effective solution.

11:30

1aNCd9. A simple estimation of z vibration for a Hopsca above depot by analog measurement and analysis in a metro station. Jiping Zhang (Environ. Impact Assessment Ctr., Zhejiang Res. and Design Inst. of Environ. Protection, 109 Tian Mu Shan Rd., Hangzhou 310007, China, jpzhang@mail.hz.zj.cn), Jianxiong Weng, Lei Wang, and Yexiang Jiang (Hangzhou Metro Group Ltd., Hangzhou 310000, China)

The word Hopsca means Hotel + Office + Park + Shopping mall + Convention + Apartment and was originated from La Defense in Paris of France in 1986. Metro depot is generally a parking and maintenance base and covers a large area and available land resources. It is a feasible approach to achieve comprehensive land utilization to develop top head estates above the depot by building residential, business, commercial, and recreational and sports urban functions. Such cases are well done, for example, in Hong Kong but generally few. The Metro in Hangzhou City of China is developing such a Hopsca. Because the trains will run and receive maintenance in the depot, the vibration and noise must be designed not transmitted to the top head estates. It will be very valuable to study the transmission and control of the vibration. This paper managed to estimate the complex vibration by simple analog measurement and analysis on a typical aboveground metro station with top head urban functions in Shanghai City of China.

11:45

1aNCd10. Acoustics of energy efficient elevator systems. Darshit Joshi (Polysonics Corp., 405 Belle Air Ln., Warrenton, VA 20186, darshitj@polysonics.com)

Traditional systems for vertical transportation require both machinery and space to operate. New, high performance, elevators eliminate these requirements, and in doing so significantly decrease the noise and vibration impacts associated with the traditional system architecture. Furthermore, machine roomless elevator systems offer superior electrical efficiency, a reduction in shaft wall requirements, centralized structural loading, and eliminate the need for pits, sumps, and the possibility for water contamination from spilled hydraulic oil. Also, interior environmental considerations are impacted by a number of innovations that result in a reduction in hydraulic pump, pump room and plumbing noise, motor and cab vibrations, and other significant improvements. Finally, the cost associated with ongoing maintenance and support such as lubrication of steel ropes is greatly reduced through the use of systems and technologies specifically engineered to reduce the need for many routine services that conventional systems require.

Session 1aNCe

NOISE-CON, Noise, and Psychological and Physiological Acoustics: Sound Quality

Wade R. Bray, Chair
HEAD acoustics, Inc., 6964 Kensington Rd., Brighton, MI 48116-8334

Contributed Papers

9:15

1aNCe1. An examination of the combined effects of loudness, tonalness, and roughness on annoyance ratings of aircraft noise. Shashikant More and Patricia Davies (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907, shashi@purdue.edu)

This study is part of a broader investigation into how noise characteristics other than loudness affect people's responses to aircraft noise. In previous studies, some evidence of increase in annoyance ratings with increased tonalness and roughness was found when loudness was kept constant and only tonalness or roughness was increased. The objective of this study was to investigate the combined effects of loudness, tonalness, and roughness on annoyance ratings of aircraft noise. A simulation program was used to generate two sets of stimuli based on flyover after take-off events of an Airbus-310, a Boeing-757, and an MD-80 aircraft. For nine stimuli in Set A based on an Airbus-310 recording, both tonalness and roughness varied while loudness kept constant. The 12 stimuli in Set B are based on recordings of all three aircrafts. In this test loudness, tonalness, and roughness were all varied. Subjects rated these sounds on an annoyance scale. Most of the subjects found to be very sensitive to tonalness when both tonalness and roughness varied and loudness was kept constant. Loudness was found to be the major contributor to annoyance when loudness, tonalness, and roughness varied simultaneously but tonalness and roughness also influenced the annoyance ratings.

9:30

1aNCe2. Determining sound quality judgments for metric development. André Fiebig and Klaus Genuit (HEAD Acoust. GmbH, Ebertstr. 30a, 52134 Herzogenrath, Germany, andre.fiebig@head-acoustics.de)

The sound quality of various technical products has become important for manufacturers. It is interpreted as among the most relevant factors regarding perceived product quality and is important in gaining market advantage. Usually, the purpose of sound quality studies is to derive a relationship between the subjective attributes and physical properties of sound. But how can valuable and meaningful subjective data be obtained? Identifying the most important psycho-physical quantities reflecting human responses to specific evaluation criteria depends on the applied method and the test design. Weighting factors for individual parameters in a combined metric depend strongly on the sounds used, the test environment, and the evaluation task. This paper will contrast the strengths and weaknesses of different methods, and discuss their level of inaccuracy in the intended application. The possibility of combining methods could be a step forward to achieve more grounded results and to enhance the explanatory power of developed metrics. Results of a few case studies will be presented, and options for combining different methods will be discussed.

9:45

1aNCe3. The need for a "green" sound of e-cars—The challenge of passung. Brigitte Schulte-Fortkamp (Inst. of Fluid Mech. and Eng. Acoust., Technische Universität Berlin, Einsteinufer 25, D-10587, Berlin, Germany, brigitte.schulte-fortkamp@tu-berlin.de)

The climate change is forcing to develop E-cars and also rethink how an E-car should sound. There is little agreement over what the sound should be. Some car manufactures seem to think about to make the cars more loud. But the question is how should a sound "sound" to provide the message about this ecological car? For sure there is a lack of collaboration to be filled amongst engineering and further future orientated science. Moreover, there is a clear need to get to know how the future client will accept such the new cars and how an appropriate sound will support this acceptation. Whenever the acoustical design of vehicles is under scrutiny, the crucial question about the appropriate method of evaluation arises. Many studies show that not only acoustic but also non-acoustic parameters have a major influence. Methods that give the opportunity to test the quality of the given ambience and to register the effects and evaluations in their functional interdependence as well as the influence of personal and contextual factors are needed to solve the upcoming problems to design the sounds of E-cars ecologically appropriate. This concept will be introduced.

10:00

1aNCe4. Prediction of subjective annoyance for transient sounds using a loudness-based impulsiveness measure. Andrew Willemsen, Mohan Rao (Dept. of Mech. Eng.-Eng. Mech., Michigan Technolog. Univ., 1400 Townsend Dr., Houghton, MI 49931, amwillem@mtu.edu), and Jasper Wong (Xerox Corp.)

This paper presents a study on the characterization of the sound quality of transient sounds via fundamental psychoacoustic measures. Specifically, the overall subjective perception of annoyance for transient sounds was studied. Through magnitude estimation and paired comparison jury evaluation experiments, the subjective annoyance magnitudes of 15 transient sounds were determined. For each sound, several objective psychoacoustic measures were calculated, and using simple linear regression models, the relationships between these objective measures and the subjective annoyance magnitudes were investigated. Examined psychoacoustic measures included loudness, sharpness, roughness, fluctuation strength, tonality, and a new loudness-based measure of impulsiveness. The new impulsiveness measure is based on the summation of the magnitudes of impulse-induced peaks in the loudness time history for a sound (calculated according to DIN 45631 /A1). The models were analyzed using several statistical measures of model significance and fit. It was found that for the transient sounds studied, significant relationships existed between subjective annoyance and each of the following psychoacoustic measures: loudness, sharpness, roughness, and loudness-based impulsiveness. These four measures were then combined into a single model for predicting subjective annoyance using multiple linear regression analysis. It was found that this model was highly correlated to the subjective annoyance of transient sounds.

Session 1aNCf

NOISE-CON, Noise, and Structural Acoustics and Vibration: Vibration Damping for Noise Control

Marshall W. Downing, Chair
Lord Corp., 2000 W. Grandview, Erie, PA 16514-0038

Contributed Papers

10:30

1aNCf1. Effect of damping on noise and vibration response of tubes. Avinash Patil (Material Sci. Corp., 6855 Commerce Blvd., Canton, MI 48187, avinash.patil@matsci.com)

Noise and vibration response of cylindrical tubes is critical in many applications such as automotive intake and exhaust manifolds, appliance tubes, and fluid flow pipes. One of the approaches of controlling noise and vibration response is to introduce damping in the tube. An investigation of tubes with dimensions similar to that of intake manifold tube of a large truck is presented in this paper. Two tubes were investigated, viz., monolithic steel tube and laminated steel tube. Dispersion relation for the tube was analyzed. Effects of structural waves propagating in axial and circumferential directions were studied for their coupling with acoustic space. Effects of ring frequency of the cylinder and critical frequency of the material were analyzed. An experimental investigation was conducted for the tubes. Experimental results are explained with the help of theoretical analysis of the tube.

10:45

1aNCf2. A comparison of displacement and velocity proportional dampers for use in aircraft. Mike Hudik and Mark Downing (Lord Corp., 2000 W Grandview Blvd., Erie, PA 16509, mike.hudik@lord.com)

Surface effect dampers have been used in aircraft to aid in the attenuation of shimmy conditions on the nose landing gear. Shimmy conditions can lead to increased cockpit noise, increased component wear on the nose landing gear, and a decrease in aircraft control during operation. Hydraulic fluid filled dampers have traditionally been used to attenuate shimmy. They are, however, susceptible to leaking. This in turn relates to a decrease in performance and ultimately an increase in maintenance cost. Surface effect dampers use metal sliding over a rubber piston to create a combination of viscoelastic and Coulomb damping. This construction is immune to leaking and offers unique performance in comparison to hydraulic. This paper provides an overview of the difference between surface effect and hydraulic dampers for shimmy attenuation. Points to be touched on will include damper principles, equations of motion, the theoretical hysteresis curve, volumetric efficiency, and observations related to product life. Part test data will be compared to theory. Furthermore, this paper will discuss the process of completing the STC for a particular aircraft, discussing both the laboratory and the on-wing testing required to show the functional characteristics are satisfied.

11:00

1aNCf3. Rail and wheel vibration absorber performance testing. James Tuman Nelson (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618, jnelson@wiai.com)

Rail and wheel dynamic vibration absorbers have been proposed as noise reduction measures for rail transit systems in Europe and the United States. These absorbers are essentially tuned spring-mass systems that add damping to the rail or wheel, and can effectively control wheel squeal. Their noise reduction effectiveness for rolling noise and rail corrugation noise is less clear. The extensive test results of under-car and wayside noise reductions are presented and discussed for wheel vibration absorbers at the San Francisco Bay Area Rapid Transit (BART) system for operations on resilient direct fixation fasteners. These results are compared with test results obtained under TCRP Project C3 at the Portland Tri-Met light rail system for both rail and wheel vibration absorbers. The reduction in pinned-pinned mode rail vibration by application of rail vibration absorbers on ballast-and-tie track at Tri-Met is discussed.

11:15

1aNCf4. The effects of structural damping on the radiated noise of circular saw blades during operation. Alan Hufnagel (Material Sci. Corp., 6855 Commerce Blvd., Canton, MI 48187, alan.hufnagel@matsci.com)

Noise from circular saw blades is a common industrial noise problem. A circular saw blade using constrained layer damping technology can significantly reduce the noise emitted during the cutting process. The noise reduction benefit of a constrained layer blade will be demonstrated. The influence of constrained layer damper design parameters on the blade noise reduction will be presented.

11:30

1aNCf5. Design and characterization of a tuned vibration absorber for exhaust components. Keith Ptak (Lord Corp., 2000 W Grandview Blvd., Erie, PA 16509, keith_ptak@lord.com)

Truck exhaust components must withstand vibration inputs from the diesel engine as well as from road inputs. Cumulative damage from the significant mileage these vehicles are exposed to can be severe; however, customers have an expectation that the components will last the life of the vehicle. Tuned vibration absorbers (TVAs) are an effective way to mitigate vibrations resulting from excitation of resonances due to random vibration exposure. A pipe section was tested to verify its expected mode shapes to aid in proper TVA placement. Subsequent tests were performed not only with TVAs but also with an equivalent mass in order to determine the true improvement due to the TVA effect. This paper covers a case study demonstrating the use of a TVA to achieve a 60% reduction in vibration.

11:45

1aNCf6. Predictability of field airborne noise isolation from laboratory testing. John LoVerde and Wayland Dong (1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

The relationship between laboratory and field noise isolation tests of nominally identical assemblies is a longstanding question in acoustics [LoVerde and Dong, J. Acoust. Soc Am. **122**, 2955 (2007)]. Many building codes allow field airborne noise isolation ratings (ASTC, NIC, or NNIC) to be five points lower than the corresponding laboratory rating (STC), but the origin of this relationship is not documented thoroughly in the literature. Different wood framed floor/ceiling systems were tested multiple times in the laboratory to determine performance prior to implementation in the field. Field testing was performed in a variety of completed buildings to determine the actual acoustical performance of the assemblies that were installed. The results of this testing for airborne sound isolation is presented. The authors will present analysis of the predictability of field test results based on laboratory test and the apparent translation from laboratory to field testing.

Session 1aNSa

Noise, Physical Acoustics, and INCE: Rocket Noise Environments I

Kent L. Gee, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

R. Jeremy Kenny, Cochair
NASA Marshall Space Flight Center, Huntsville, AL 35812

Jared M. Haynes, Cochair
NASA Marshall Space Flight Center, Huntsville, AL 35812

Chair's Introduction—7:55

Invited Papers

8:00

1aNSa1. Perspective on launch noise: Measurement, prediction, and characterization. Sally Anne McInerny (Dept of Mech. Eng., Univ. of Alabama at Birmingham, BEC 356D, 1530 3rd Ave S., Birmingham, AL 35294)

The noise generated by launch vehicles is examined in terms of the key source, radiation, instrumentation, and estimation issues. Basic source and acoustic radiation characteristics of un-deflected rocket exhausts are compared and contrasted with those of supersonic jets. Next, the instrumentation demands of, and unique characterization methods applied to, rocket noise sound pressure measurements are discussed. Finally, on vehicle sound pressure measurements during the launch from a covered pad are presented. Contradictions between these loads and those predicted using traditional lift-off loads methods are highlighted. This paper concludes with a brief outline of research and instrumentation needed to improve agreement between the physics of the prediction model and the actual measured lift-off loads.

8:40

1aNSa2. Liftoff acoustic environment of the sounding rocket ATK Launch Vehicle (ALV-X1): Prediction versus measurement. J. Houston (Jacobs/ESTS Group, 1525 Perimeter Parkway, Ste. 330, Huntsville, AL 35806), Douglas Counter, R. Jeremy Kenny (NASA Marshall Space Flight Ctr., Huntsville, AL), and John Murphy (ATK Space Systems, Elkton, MD)

Launched from the Mid-Atlantic Regional Spaceport (MARS) Pad 01B on August 22, 2008, the ATK Launch Vehicle (ALV-X1) provided an opportunity to measure liftoff acoustic noise data. Predicted liftoff acoustic environments were developed by both NASA MSFC and ATK engineers. ATK engineers developed predictions for use in determining vibro-acoustic loads using the method described in the monograph NASA SP-8072. The MSFC ALV-X1 liftoff acoustic prediction was made with the Vehicle Acoustic Environment Prediction Program (VAEPP). The VAEPP and SP-8072 methods predict acoustic pressures of rocket systems generally scaled to existing rocket motor data based upon designed motor or engine characteristics. The predicted acoustic pressures are sound-pressure spectra at specific positions on the vehicle. This paper presents the measured liftoff acoustics on the vehicle and tower. Additionally, the ALV-X1 liftoff data can be scaled to define liftoff environments for the NASA Constellation program Ares vehicles.

9:00

1aNSa3. Acoustic design of launch pad for advanced solid rocket. Seiji Tsutsumi, Kota Fukuda, Ryoji Takaki, Tatsuya Ishii, and Kyoichi Ui (JAXA, 3-1-1 Yoshinodai Sagamihara, Kanagawa 229-8510, Japan)

In the mission of the Advanced Solid Rocket studied in JAXA, the launch-pad is required to improve the operation performance as well as to reduce the cost. While, exhaust plume from the solid booster generates severe acoutic wave so that decrease in the acoustic level at lift-off is also an important design issue. Preliminary trade-off analysis is peformend by using computational fluid dynamics. Major noise sources such as the Mach wave and the impingement noise and their correlation with the vehicle's altitude are revealed. The knowledge of the acoustic characteristics gives us idea how to decrease the acoustic level within the mission requirements. Based on the configuration found in the numerical study, subscale test using 1/43 scale mock-up is planned to ensure the acoustic level around the vehicle.

9:20

1aNSa4. Prediction of the acoustic environment on the launch pad for the Ares I rocket, including the effects of drift and water suppression. Kenneth J. Plotkin (Wyle Labs., 241 18th St. S., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wyle.com)

A model, PAD, has been developed for prediction of noise in the vicinity of launch vehicles, with specific application to the mobile launcher and tower for the Ares I launch vehicle. It follows the basic principles of a traditional NASA model [NASA Report No. SP-8072 (1971)] but uses an updated rocket sound power source model developed by Sutherland [AIAA Paper No. 93-4383 (1993)] together with updates for the distribution and directivity of sources along the rocket plume. The effect of the deflector on plume geometry and shielding by the launcher deck are included in the model. Overall and spectral (octave band or psd) levels are predicted at specific points or on a regular grid of lateral and vertical positions on and around the launcher. A basic version of PAD was used for nominal acoustic predictions for vertical lift with all exhaust flow going through the flame hole in the deck and no water suppression (AIAA Paper No. 2009-3163). The model has since been extended to account for lateral drift of the vehicle, with consequent impingement of the plume on the deck and for the effect of water suppression systems. [Work supported by the National Aeronautics and Space Administration.]

9:40

1aNSa5. A post-flight ignition overpressure review for the Arex I-X test rocket. David A. Alvord (1525 Perimeter Parkway, Ste. 330, Huntsville, AL 35806)

Ares I-X post flight ignition overpressure review ignition overpressure (IOP) is an unsteady fluid flow and acoustic phenomena caused by the rapid expansion of gas from the rocket nozzle within a ducted launching space resulting in an initially higher overpressure wave. This wave is potentially dangerous to the structural integrity of the vehicle. An in-depth look at the IOP environments resulting from the Ares I-X solid rocket booster configuration showed good correlation between the pre-flight predictions and post-flight analysis results. Correlation between the chamber pressure and IOP transients showed successful acoustic mitigation, containing the strongest IOP waves below the mobile launch pad deck. The flight data allowed subsequent verification and validation of Ares I-X unsteady fluid ducted launcher predictions, computational fluid dynamic models, and strong correlation to historical shuttle data.

10:00—10:20 Break

10:20

1aNSa6. High-speed jet noise: A two source model. Christopher K. W. Tam (Dept. of Mathematics, Florida State Univ., Tallahassee, FL 32306-4510, tam@math.fsu.edu)

Recent experiments on high-speed jets indicate that jet noise is made up of two components. One component dominates in a conical sector surrounding the jet axis in the downstream direction. Experimental evidence suggests that this noise component is generated by the large turbulence structures of the jet flow. The other component is dominant in the sideline and upstream directions. There is ample experimental evidence showing that this noise component is generated by the fine scale turbulence of the jet. Optical observations confirm that there is a separation of turbulence scales in jets. Jet turbulence consists of large structures with length scale of the order of the jet diameter and fine scale turbulence with a much smaller length scale. These two types of turbulent fluid motions, with large disparate length scales, co-exist in the first one and half core length of the jet. In this presentation, extensive far field noise and direct correlation data are provided to support the two-noise source model. The validity of the two-noise source model is also strongly supported by near acoustic field measurements.

10:40

1aNSa7. Near-field energy-based measurements of small solid rocket motors. Kent L. Gee, Jarom H. Giraud (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 Eyring Sci. Ctr., Provo, UT 84602, kentgee@byu.edu), and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Results of detailed near-field energy-based measurements of small solid rocket motors are described. The 5-in. center perforated motors are tested to ensure thrust equivalence of Shuttle reusable solid rocket motor propellant samples. An array of four three-dimensional acoustic probes, located as close as two nozzle diameters to the plume shear layer, was located at nine positions for a total of 36 measurement points. The data collected were used to create maps of near-field energy-based quantities in the vicinity of the plume. In particular, the time-averaged vector intensity reveals the direction of the net acoustic energy flow away from the rocket plume, permitting direct characterization of the source region from these motors as a function of frequency. Further, these measurements establish the utility of energy-based acoustic quantities in aeroacoustic source measurement and modeling. [Work supported by an STTR from NASA Stennis Space Center. The cooperation of ATK Space Systems Test Services is gratefully acknowledged.]

11:00

1aNSa8. Acoustic measurements for small solid rocket motors. Magda B. Vargas (All Points Logistics/ESTS Group, NASA Marshall Space Flight Ctr., Huntsville, AL 35812, magda.b.vargas@nasa.gov) and R. Jeremy Kenny (NASA Marshall Space Flight Ctr., Huntsville, AL 35812)

Rocket acoustic noise can induce loads and vibration on the vehicle as well as the surrounding structures. Models have been developed to predict these acoustic loads based on scaling existing solid rocket motor data. The NASA Marshall Space Flight Center acoustics team has measured several small solid rocket motors (thrust below 150 000 lbf) to anchor prediction models. These data will provide NASA the capability to predict the acoustic environments and consequent vibro-acoustic response of larger rockets (thrust above 1 000 000 lbf) such as those planned for the NASA Constellation program. This paper presents the methods used to measure acoustic data during the static firing of small solid rocket motors and the trends found in the data.

11:20

1aNSa9. An acoustical comparison of sub-scale and full-scale. Jared M. Haynes and Robert J. Kenny (Marshall Space Flight Ctr., MS ER42, Huntsville, AL 35812)

Recently, members of the Marshall Space Flight Center (MSFC) Fluid Dynamics Branch and Wyle Labs measured far-field acoustic data during a series of three reusable solid rocket motor (RSRM) horizontal static tests conducted in Promontory, UT. The test motors included the Technical Evaluation Motor 13 (TEM-13), Flight Verification Motor 2 (FVM-2), and the Flight Simulation Motor 15 (FSM-15). Similar far-field data were collected during horizontal static tests of sub-scale solid rocket motors at MSFC. Far-field acoustical measurements were taken at multiple angles within a circular array centered about the nozzle exit plane, each positioned at a radial distance of 80 nozzle-exit-diameters from the nozzle. This type of measurement configuration is useful for calculating rocket noise characteristics such as those outlined in the NASA SP-8072 "Acoustic Loads Generated by the Propulsion System." Acoustical scaling comparisons are made between the test motors, with particular interest in the overall sound power, acoustic efficiency, non-dimensional relative sound power spectrum, and directivity. Since most empirical data in the NASA SP-8072 methodology are derived from small rockets, this investigation provides an opportunity to check the data collapse between a sub-scale and full-scale rocket motor.

MONDAY MORNING, 19 APRIL 2010 GRAND BALLROOM V, 9:30 TO 11:50 A.M.

Session 1aNSb

Noise and Architectural Acoustics: Soundscape Concert I

Alex U. Case, Cochair
Fermata Audio & Acoustics, P.O. Box 1161, Portsmouth, NH 03802-1161

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Einsteinufer 25, Secr TA 7, Berlin, 10587, Germany

Invited Papers

9:30

1aNSb1. Multichannel soundscape recording techniques. Alex U. Case (Sound Recording Tech., UMass Lowell, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

Several goals motivate the recording a soundscape: measurement, archiving, mitigation, preservation, design, modeling, education, marketing, and entertainment. Salient properties of a soundscape are similarly broad: levels versus frequency and time, their averages and associated statistics, discrete sound source identities, distance, localization, size, spaciousness, envelopment, and additional subjective sound qualities, sometimes subtle. While traditional noise metrics are based on single microphone measurements, a soundscape is better captured, quantified, and studied using multi-microphone techniques. Multichannel recording techniques for large-scale soundscapes, borrowed from the multitrack recording practices of concert halls and recording studios, are defined and compared. The resulting data and audio files are evaluated for their ability to satisfy the needs of owners, designers, and end-users of a soundscape.

10:10—10:30 Break

10:30

1aNSb2. Listen to the change of a soundscape. Klaus Genuit and Andre Fiebig (HEAD acoustics GmbH, Ebertstr. 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

The general understanding of soundscapes as musical compositions requires the analysis of the sonic environment as a whole as well as in its different facets comparable to the study of music. However, soundscape studies often differ in the way the investigated soundscapes are recorded. In addition, a lot of soundscape articles lack an exhaustive description of the measurement conditions and protocol. This impedes the comparison of soundscape studies and their results. Therefore, it seems there is a need for measurement guidelines to provide a common basis allowing for subsequent comparative studies. The authors have proposed in previous papers measurement guidelines which could help to establish a common ground with respect to soundscape measurements. To check the adequacy and applicability of the recommended procedures for the (physical) measurement of soundscapes, several recordings of historic places in Aachen are realized and analyzed. Here, binaural recordings are performed, which allows for an aurally-accurate reproduction of the environmental sound. In supplemental laboratory tests, the conflict between the experience of the sound in the real urban space and the listening in artificial test situations is examined. Furthermore, the paper presents the change of the soundscape and its composition over time.

11:10

1aNSb3. Cultural soundscape of the Grant-Kohrs Ranch national historic site. Robert C. Maher (Elec. and Comput. Eng., Montana State Univ., 610 Cobleigh Hall, Bozeman, MT 59717-3780, rob.maher@montana.edu)

It is often noted that although many parks and historic sites *look* as they did a century ago, very few *sound* as they once did. Management principles of U.S. National Parks and Historic Sites include the need to "preserve soundscape resources and values of the parks to the greatest extent possible to protect opportunities for appropriate transmission of cultural and historic sounds that are fundamental components of the purposes and values for which the parks were established." Therefore, a critical part of soundscape management is to determine the existing ambient soundscape to detect trends and changes. An ongoing project at Grant-Kohrs Ranch National Historic Site, located just north of Deer Lodge, Montana, involves a year-long continuous audio recording and soundscape characterization. The recording status and example soundscape excerpts are presented and discussed.

MONDAY MORNING, 19 APRIL 2010 HARBORSIDE B, 8:00 A.M. TO 12:00 NOON

Session 1aPP

Psychological and Physiological Acoustics and Noise: Noise-Induced Hearing Loss: From Physiology to Prevention

Sharon G. Kujawa, Cochair
Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114

Kim S. Schairer, Cochair
Univ. of Wisconsin, Dept. of Communicative Disorders, 1975 Willow Dr., Madison, WI 53706

Chair's Introduction—8:00

Invited Papers

8:05

1aPP1. Epidemiology of noise-induced hearing loss. Yuri Agrawal (Dept. of Otolaryngol.-Head and Neck Surgery, Johns Hopkins Univ., 601 N. Caroline St., Baltimore, MD 21287)

Noise exposure is a significant risk factor for hearing loss; recent analyses from a national sample of US adults have shown that occupational and firearm noise exposure increase the risk of hearing loss by 60% and 90%, respectively. Noise trauma typically produces a high-frequency pattern of hearing loss due to hair cell injury in the cochlear base, in contrast to cardiovascular risk factors which appear to cause cochlear damage across the frequency range. Significant interactions between noise exposure and cardiovascular risks have been demonstrated, such that these exposures exert a multiplicative detrimental effect on hearing thresholds. Although exposure to industrial occupational noise may be decreasing in the US population, the rising use of personal listening devices—particularly among children and young adults—is raising concern for a surge in hearing loss prevalence in the younger population. Indeed, recent data from the US population suggest that during the 1999–2004 time interval, significant increases in the prevalence of hearing loss were observed only among young adults. Further longitudinal analyses of children and young adults will be required to better characterize the hearing risks associated with these new forms of noise exposure.

8:40

1aPP2. Biology of noise-induced hearing loss. M. Charles Liberman (Eaton-Peabody Labs., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114)

This talk will review recent research on the cellular and molecular changes underlying noise-induced hearing loss (NIHL) and potential therapeutic approaches to its prevention and/or mitigation. Topics will include (1) the functionally important structural changes underlying reversible versus irreversible NIHL, (2) the factors affecting vulnerability to NIHL, (3) the inadequacy of threshold testing as a measure of noise-induced cochlear damage, and (4) the possible utility of antioxidant and other chemical therapies.

9:15

1aPP3. Spectrotemporal integration in listeners with normal hearing and those with noise induced hearing loss: An application of the Meddis Matlab Auditory Periphery (MAP) model. Lawrence L. Feth, Evelyn M. Hoglund, Yonghee Oh (Dept. of Speech and Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210), and Ray Meddis (Univ. of Essex, Colchester CO4 3SQ, United Kingdom)

Our recent study of spectrotemporal integration by human listeners with normal hearing has shown that the detection thresholds for signals made up of multiple bursts of brief tones improve as the number of bursts is increased. Elementary signals consisted of 10-ms tone bursts centered on 1 of 12 frequencies spaced 1 ERB apart. Complexes were constructed by adding individual tone bursts along the temporal or spectral dimension, or both. Quiet thresholds were measured using a one interval adaptive procedure. There was no significant difference in performance for a variety of time by frequency patterns used to construct the signals. Even random selection of tone frequencies for each signal presentation did not degrade integration performance. The current study measured spectral integration, temporal integration, and spectrotemporal integration in listeners with noise induced hearing loss as well as in normal-hearing listeners. The Meddis MAP model was adjusted to fit integration performance for the normal-hearing listeners and then used to predict the performance by NIHL listeners by modifying the outer hair cell parameters. [Research supported by a grant from the Office of Naval Research # N000140911017.]

9:30

1aPP4. Effect of noise-induced hearing loss on a multi-channel listening task. Kim Abouchacra (Dept. of Otolaryngol.-Head and Neck Surgery, Medical Ctr., AUBMC, American Univ. of Beirut, P.O. Box 11-0236, Beirut, Lebanon, ks05@aub.edu.lb), Janet Koehnke, Joan Besing (Montclair State Univ., Montclair, NJ 07043), and Tomasz Letowski (U.S. Army Res. Lab., Aberdeen Proving Ground, MD 21005-5425)

Monitoring multi-channel radio communication is a common activity for many military and civilian professionals, such as those who work in tactical operation centers, command-control towers, and voice interception facilities. In such environments, communication is critical and errors in speech understanding could cost time, equipment, and even loss of life. When an individual has a hearing loss, these costs have the potential to increase substantially. The purpose of this study was to compare speech recognition scores (SRSs) of listeners with normal hearing and noise-induced hearing loss, when target messages were presented with two, three, or four interfering messages through various spatial and non-spatial auditory displays. For both groups, SRSs decreased for all display types as the number of competing messages increased. However, listeners with hearing impairment had significantly poorer SRSs than listeners with normal hearing, with decreases in scores ranging from 10%–25% in all conditions. While both groups benefited from spatial auditory displays, listeners with hearing impairment demonstrated significantly greater improvement in SRS than those with normal hearing. These findings support the use of spatial displays for multi-channel monitoring tasks and highlight potential costs associated with hearing loss especially in the absence of spatial displays.

9:45—9:55 Break

9:55

1aPP5. Behavioral consequences of weakened medial olivocochlear efferent noise-control mechanisms in mice. Amanda M. Lauer and Bradford J. May (Dept. of Otolaryngol.-HNS, Johns Hopkins Univ., 521 Traylor Bldg., 720 Rutland Ave., Baltimore, MD 21205)

The mammalian medial olivocochlear (MOC) efferent feedback system protects the ear from intense noise exposure, aids in cochlear gain control, and improves responses to signals in noise. Unequivocal behavioral effects of eliminating MOC feedback have been difficult to demonstrate. We used an acoustic startle reflex (ASR) modification paradigm to reveal behavioral effects of eliminating MOC feedback in mice. Mice lacking the alpha9 nicotinic acetylcholine receptor subunit showed increased facilitation of the ASR reflex in the presence of background noise, abnormal responses to brief fluctuations in background noise, and abnormal inhibition of the ASR by changes in the location of noise in azimuth and elevation. Abnormal behavioral responses occurred in the absence of cochlear hearing loss. The behavioral effects of weakened MOC feedback may reflect both direct and indirect effects of abnormal cochlear gain control in the presence of noise. The results provide further evidence for a role of MOC efferent feedback in hearing in noise and demonstrate that noise-related hearing deficits can occur when hearing thresholds are normal.

10:10

1aPP6. Susceptibility to intense impulse noise: Evidence from the Albuquerque dataset. G. Richard Price (Auditory Hazard Anal., P.O. Box 368, Charlestown, MD 21914, Ahanal.@comcast.net)

For theoretical and applied reasons, there has long been an interest in the problem of susceptibility to hearing loss from intense sounds; yet no useful predictive indices have been found. Some evidence suggests that susceptibility cannot be considered as a constant for individual ears; however, thinking and experimental designs often assume that it is. A critical data set in this regard are the US Army's Albuquerque studies [Johnson, D. L. (1994) USAARL Contract Report No. 94-2, U.S. Army Aeromedical Research Laboratory, Rucker, AL] in which human Ss were exposed repeatedly to explosive sources. An analysis of individual threshold shift data finds that of 28 instances in which ears showed a threshold shift 15 dB or higher, 25 Ss subsequently passed higher levels of exposure, sometimes much higher levels. This outcome was clearly contrary to the expectation that they were susceptible ears. Issues such as HPD fit and random elements in exposures are demonstrably a part of the problem. Susceptibility, at present, should probably be considered a statistical concept, appropriate to groups of Ss.

Invited Paper

10:25

1aPP7. Noise-induced hearing loss in the U.S. Military: A review. Larry Humes (Dept. of Speech & Hearing Sci., Indiana Univ., Bloomington, IN 47405, humes@indiana.edu)

This presentation will review some of the key findings of an Institute of Medicine committee report [Humes *et al.* Noise and Military Service: Implications for Hearing Loss and Tinnitus (The National Academies, Washington, DC, 2005.] on noise-induced hearing loss in U.S. military personnel since World War II. With regard to hearing loss, the majority of the data available at the time of the report were average group hearing thresholds from cross-sectional studies. Patterns of hearing loss consistent with noise exposure could be seen in several cross-sectional studies of military personnel. These average data indicated that hearing thresholds were worse in those groups with more years of military service. However, these cross-sectional data were not a sufficient basis for attributing greater hearing loss solely to a longer exposure to military noise while in the military. With the information available, it was not possible to establish the precise proportion of a given military population that developed noise-induced hearing loss during military service, the amount of hearing loss incurred, or the relative risk of noise-induced hearing loss for a given individual, based on his or her branch of military service, occupational specialty, or service era.

Contributed Paper

11:00

1aPP8. National occupational research agenda for hearing loss in the manufacturing sector. William J. Murphy (CDC/NIOSH Hearing Loss Prevention Team, 4676 Columbia Parkway, MS C-27, Cincinnati, OH 45226-1998)

The National Institute for Occupational Safety and Health (NIOSH) has developed the National Occupational Research Agenda (NORA) as a means of organizing and prioritizing its research efforts in occupational safety and health. Research has been categorized across 8 industrial sectors and 22 cross-sectors. Among these, the manufacturing sector is the largest sector in which hearing loss occurs due to occupational exposures. The NIOSH Manufacturing Sector Council has developed strategic research goals to be completed by government, academic, and industry stakeholders. Foremost among these goals were the identification and inventory of sources of noise exposure resulting in hearing loss, and the determination of behavioral and exposure factors that contribute to the risk of hearing loss. This paper will outline the efforts of the Manufacturing Sector Council to identify research needs, outline research partnership opportunities, and seek effective means of implementing the research agenda to reduce the incidence of hearing loss among American workers.

Invited Paper

11:15

1aPP9. Development of therapeutics to protect the inner ear: Support from animal models. Colleen Le Prell (Dept. of Communicative Disord., Univ. of Florida, Gainesville, FL colleeng@phhp.ufl.edu), Kevin Ohlemiller (Central Inst. for the Deaf, Washington Univ., St. Louis, MO), David Dolan, and Josef Miller (Univ. of Michigan, Ann Arbor, MI)

Noise-induced hearing loss (NIHL) is a significant clinical, social, and economic issue. Although we once thought virtually all NIHL was a consequence of mechanical damage to cells in the inner ear, we now know that intense metabolic activity drives the formation of free radicals (short-lived, unstable, highly reactive clusters of atoms) in the inner ear. Studies in animals have clearly shown that free radicals formed during and after noise importantly contribute to NIHL, and many laboratories have demonstrated that free radical scavengers ("antioxidants") reduce NIHL in animal subjects. Our group, including investigators at multiple institutions, has specifically shown that the combination of beta-carotene, vitamins C and E, and magnesium is highly effective in preventing NIHL and sensory cell death in guinea pigs and mice. Use of free radical scavengers, or antioxidants, to prevent noise-induced deficits has thus become a clinical research goal. Given the multitude of intracellular events that occur during and after noise, it is not surprising that other agents are also potentially useful therapeutic agents. Translational investigations are essential to confirm potential utility of these agents in the human inner ear.

11:50—12:00 Summary/Discussion

Session 1aUW

Underwater Acoustics: Acoustic Imaging and Target Detection

Zachary J. Waters, Chair
Naval Research Lab., Code 7130, Washington, DC 20345-5320

Contributed Papers

8:00

1aUW1. Laboratory experimental results for synthetic aperture sensing applied to mine detection. S. C. Walker, W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093), Stephane Aumary, and Ghislain Marais (French Naval Acad. Res. Inst. (IRENAV))

Recent interest in mine detection has focused on the use of synthetic aperture sonar (SAS) array processing techniques to improve the along track sensor resolution of mobile side looking sonar platforms. We present the results of a series of laboratory experiments designed to study the effectiveness of various SAS array processing techniques for detection, localization, and classification of proud and buried objects. The experiments are carried out on a small spatial scale at ultrasonic frequencies. Notably, we find it possible to detect, localize, and classify buried objects (on the order of a wavelength in scale) from a great distance (hundreds of wavelengths).

8:15

1aUW2. Characterization and modeling of synthetic aperture sonar imagery. Anthony P. Lyons (Appl. Res. Lab., Penn State Univ., State College, PA 16804, apl2@psu.edu), Douglas A. Abraham (CausaSci LLC, Arlington, VA 22205), and Shawn F. Johnson (Johns Hopkins Univ., Laurel, MD 20723)

The characterization and modeling of synthetic aperture sonar (SAS) image statistics are of importance for developing target-on-background detection and classification algorithms and for developing specialized filters for speckle noise reduction. A simple model is presented to predict the impact of amplitude scaling caused by seafloor ripples on SAS image speckle statistics. The continuous variation in scattering strength produced by ripples (i.e., ripple-induced changes in seafloor slope) is treated as a deterministic amplitude scaling on image speckle produced by the SAS imaging process. Changes in image statistics caused by ripples are quantified in terms of an effective K-distribution shape parameter. Agreement between shape parameter estimated from the scaling model and from SAS data collected in experiments off of Panama City, FL and off of the Ligurian coast of near La Spezia, Italy illustrates the efficacy of the model. [Work supported by ONR Grant Nos. N00014-04-1-0013 and N00014-06-1-0245.]

8:30

1aUW3. Time-domain numerical simulation of scattering by objects in the seafloor. David C. Calvo (Acoust. Div., Naval Res. Lab., Washington, DC 20375), Mario Zampolli (TNO Defense, Security and Safety, The Hague, Netherlands), Jill P. Bingham, Saikat Dey, and Harry J. Simpson (Naval Res. Lab., Washington, DC 20375)

Numerical simulation of scattering in the time domain offers potential implementation advantages in comparison with frequency-domain methods. By using explicit numerical time integration schemes, elastic stresses and velocities can typically be advanced using only spatially local information which makes parallelization by domain decomposition straightforward to implement in comparison with the solution of a large linear system of equations. On the other hand, incorporation of frequency-dependent sound speed and attenuation appears more challenging in the time domain with the general need to compute convolutions under non-FFT friendly conditions. This talk presents efficiency gains for the elastodynamic finite integration technique (EFIT) which is similar to the finite-difference time-domain method (FDTD). We present results on using explicit decomposition of scattered and incident wave fields and far-field projection in the time domain with the discrete Helmholtz–Kirchhoff integral. We also present incorporation of frequency-dependent material parameters by efficient recursive convolution techniques that have been used successfully to create time-domain perfectly matched layers [J. A. Roden and S. D. Gedney, Microwave Opt. Technol. Lett. **27**, 334–338, (2000)]. As an application example, we present results on computing bistatic target strength variance for a two-dimensional object located near a statistically rough seafloor. [Work sponsored by the Office of Naval Research.]

8:45

1aUW4. Testing and modeling of a spiral wave front beacon for use in navigation. Benjamin R. Dzikowicz (Naval Res. Lab., Code 7130, 4555 Overlook Ave., Washington, DC 20375-5320) and Brian T. Hefner (Univ. of Washington, Seattle, WA 98105-6698)

A beacon with two cylindrical elements of 3-1 composite material has been developed, tested, and modeled for use in navigation. The displacement of each element is radial and the phase of the outgoing wave front is determined by the cross-section of the cylinder. One element, the reference, has a circular cross-section which produces circular wave fronts. The second element has a radius that varies linearly with angle such that there is a displacement of one wavelength over 360 deg; the resulting wave front is spiral. This beacon improves on a prototype presented at a prior meeting [Dzikowicz and Hefner, J. Acoust. Soc. Am. **125**, 2540 (2009)]. Results from experiments conducted at the Naval Surface Warfare Center, Panama City Division are compared with finite element models. Surface displacement of the elements is determined by the use of laser Doppler vibrometry. Variation in the field due to geometry and surface displacement variation are discussed. A simple navigational test shows the efficacy of the beacon. [Work supported by the Office of Naval Research.]

9:00

1aUW5. Small scale test bed for studying multiaspect and multistatic sonar systems. Patrick C. Malvoso, John S. Stroud, Raymond Lim, Joseph L. Lopes (Naval Surface Warfare Ctr., Panama City Div. Code T11, Panama City, FL 32407, john.stroud@navy.mil), and Benjamin R. Dzikowicz (Naval Res. Lab., Washington, DC 20375-5320)

A nominally 1:50 scale acoustic test bed is operational at the Naval Surface Warfare Center-Panama City Division (NSWC-PCD) to image free-field, bottom, and buried targets using multiaspect, bistatic, backscatter, forward scatter, and synthetic aperture sonar (SAS) techniques. The test bed is designed to test and study novel geometries and techniques faster and cheaper than can be done in the field. Using precise translational positioning systems mitigates any issues associated with positioning of multiple sonar platforms in the field. The use of two transmit and receive transducer platforms is superior to a single platform, which is restricted to purely backscattered imaging, for producing SAS images. With two platforms, one can generate four images, two using backscattered return data and two using data sent from one platform to the other. Thus, the system is inherently both dual aspect and bistatic. The setup, including the scaled sediment, acoustic sources and receivers, and computer controlled translational stages, is described. Preliminary beamformed data are presented showing the relationships between each of the images described above. In addition, techniques to

mitigate the interference of direct path returns in forward-scattering geometries are discussed. [Work sponsored by the Office of Naval Research.]

9:15

1aUW6. Forward scatter target strength extraction in a littoral environment. Larry A. Kraus, Alain R. Berdoz, Danial L. Amon, Philip A. Frank, Steve W. Liskey (Global Strategies Group (North America), 2200 Defense Hwy., Crofton, MD 21114), Joseph A. Bucaro (Excet, Inc., Springfield, VA), Harry J. Simpson, Brian H. Houston, and David C. Calvo (Naval Res. Lab., Washington, DC)

A rail-based system was used to collect forward and near-forward scattered echoes from a spherical shell in 14-m waters near Shell Island, Panama City, FL. The source was positioned 25 m from the scattering target and the 48-m horizontal rail on the opposite side, also 25 m from the target. The major obstacle to obtaining high-quality forward scatter target strength versus frequency and angle is the extraction of the much stronger time and position overlapping incident source signal. In previous laboratory measurements, this is accomplished with high precision by direct measurement of the incident field before the scattering target is positioned, a method not possible in a target search scenario or in a less stable environment. Here an attempt is made to obtain the forward scattered target strength by post-processing the received signals obtained in the littoral environment which contain both the echo and the overlapping source signal. The methodology involves using a wavenumber domain filter to remove the incident wave followed by standard synthetic aperture procedures to extract the desired target scattered signal. The resulting forward scattered target strength will be compared to what we expect analytically and experimentally from this simple target. [Work supported by ONR].

9:30

1aUW7. Forward scatter and backscatter low-frequency synthetic array measurements of the structural acoustic response from proud targets using a 48-m-long rail in a littoral environment. Harry J. Simpson, Zachary J. Waters, David C. Calvo, Brian H. Houston (Physical Acoust., Branch Code 7130, Naval Res. Lab., Washington, DC 20375), Alain R. Berdoz, Danial L. Amon, Philip A. Frank, Steve W. Liskey, Larry A. Kraus (Global Strategies Group (North America), Crofton, MD), and Joseph A. Bucaro (Excet, Inc., Springfield, VA)

A series of short range (25-m) forward scatter and backscatter measurements in a littoral environment was conducted to quantify the structural acoustic response from proud targets in the 2–23 kHz frequency band. The water channel was 14 m deep in the Gulf of Mexico near Shell Island, Panama City FL. The bottom was a medium grained sand. The acoustic forward scatter response of the proud targets was measured in a bistatic configuration with the source 25 m from the target and a receiver mounted on a 48-m-long rail. The rail is used to position the receiver and synthetically quantify the structural acoustic forward scatter response. A second source was co-located with the receiver on the rail, and monostatic backscatter measurements were also taken for each target. The structural acoustic response was analyzed and will be reported. The synthetic array experimental results are compared and contrasted with laboratory measurements. The laboratory measurements are convolved with water channel propagation predictions using a parabolic equation model of the littoral environment. The convolved laboratory target strengths compare well with the synthetic array measurements, and a detailed discussion of the comparison will be presented. [Work supported by ONR.]

9:45

1aUW8. Detection of a resonant target in shallow water using iterative, single-channel time reversal at low frequencies (2–23 kHz). Zachary J. Waters, Harry J. Simpson, Benjamin R. Dzikowicz, Brian H. Houston (Physical Acoust. Branch Code 7130, Naval Res. Lab., Washington, DC 20375-5320), Alain R. Berdoz, and Larry A. Kraus (Global Strategies Group (North America), Crofton, MD 21114)

Iterative time reversal with a single-channel transducer has been shown to enhance echoes from resonant targets both in scaled high-frequency (0.5–2-MHz) laboratory experiments [Waters *et al.*, IEEE UFFC **56**, 1429 (2009)] and in mid-frequency (20–200-kHz) pond experiments [Dzikowicz *et al.*, J. Acoust. Soc. Am. 123, 3755 (2008)]. These techniques are applied to study scattered returns from a proud, spherical target at low frequencies, 2–23 kHz, in a littoral environment. Monostatic scattering experiments are conducted in the Gulf of Mexico near Panama City, FL in 14-m-deep water. The target, a 60-cm-diameter stainless steel spherical shell (1.5 cm thick), rests on a sandy bottom at a range of 35 m from a rail-mounted source-receiver pair positioned at a height of 2.3 m from the bottom. Through iterative retransmission of windowed and time reversed scattering returns, enhancement of target echoes with respect to background noise and reverberation is explored. Frequencies to which the time reversal process converges are compared with analytical predictions and pristine laboratory measurements of the free-field scattering response. [Work supported by ONR.]

10:00—10:30 Break

10:30

1aUW9. Wave separation and signal isolation via line-scan supersonic acoustic holography. Timothy M. Marston, Philip L. Marston (Dept. of. Phys. and Astronomy, Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu), and Kevin L. Williams (Univ. of Washington, Seattle, WA 98105)

Pressure signals scattered from a target and received by a line array can be buried in noise having similar frequency content, such as reflections from nearby interfaces or objects adjacent to the target. In such circumstances, it may be difficult to isolate the desired signal through the use of standard linear time-invariant filtering techniques, because the frequency content of the signal and noise are identical. Supersonic line-scan holography, however, can be used in many such circumstances to achieve signal isolation by back-propagating the acoustic signal to a focal region, spatially and temporally filtering the signal around the focal region, and forward-propagating the isolated signal to the original array location. This technique was used to measure the bistatic angular scattering pattern of a small target driven at resonance. This was done in a water tank using a hydrophone scanned along a line. A demonstration of how this method was used to isolate the signal from larger, overlapping reflections will be given. The method was also applied to backscattering measured with a scanned transducer and fixed proud cylinder resting on sediment. [Work supported by ONR.]

10:45

1aUW10. Scattering of linear frequency modulated chirps by resonators: Effect on pulse compression. Jon La Follett and Philip L. Marston (Dept. Phys. and Astronomy, Washington State Univ., Pullman, WA 99164-2814, jon.lafollett@email.wsu.edu)

Pulse compression is a common technique used to improve the signal to noise ratio and time resolution of a sonar system. A linear frequency modulated (LFM) chirp is one type of excitation used in conjunction with matched filtering to produce pulse compressed time domain data. This approach uses a correlation between the transmitted and received signals. Scattering by a target having a narrow resonance within the bandwidth of the LFM chirp may contain a superposition of specular scattering and scattering features associated with the transient response of the resonator. If the frequency sweep rate is not sufficiently slow relative to the ring-down time of the resonator, the transient response may result in artificial features and decreased time resolution in the pulse compressed signal due to decreased correlation with the transmitted signal. To explore this effect, water tank experiments were carried using a water filled cylindrical shell as the resonant target. Experiments demonstrate that for sufficiently rapid chirp rates, the target response envelope has an oscillatory modulation that affects the pulse compressed signal. [Research supported by ONR.]

11:00

1aUW11. Scattering experiments for exploring pressure gradient coupling to modes of solid cylinders. Timothy M. Marston and Philip L. Marston (Dept. Phys. and Astronomy, Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu)

Grazing illumination of a target close to a flat interface as well as evanescent wave illumination of a buried target can produce significant vertical acoustic pressure gradients at a target. A prior experiment [Espana *et al.*, J. Acoust. Soc. Am. 126, 2187 (2009)] demonstrated the selective coupling to a mode of a flat-ended solid horizontal circular cylinder. The cylinder's axis was positioned so that the vertical pressure gradient was large at the end of the cylinder. The present investigation explores this coupling for wider variety cylinders in which the end of the cylinder was modified to include a flat paddle. Such a paddle resembles a fin-like appendage. When a flat vertical paddle is subjected to a vertical pressure gradient, the associated torque on the paddle facilitates the excitation of torsional modes of the solid cylinder. The radiation damping is found to increase for larger paddles. Targets modified in this way facilitate the investigation of specific scattering mechanisms. Coupling to this type of mode is suppressed when the cylinder's axis is at a pressure anti-node where the pressure gradient vanishes. [Work supported by ONR.]

11:15

1aUW12. Boundary enhanced helical ray coupling to high-frequency modes of solid aluminum cylinders. Jon La Follett and Philip L. Marston (Dept. Phys. and Astronomy, Washington State Univ., Pullman, WA 99164-2814, jon.lafollett@email.wsu.edu)

Surface guided leaky Rayleigh waves have been shown to provide a significant contribution to the backscattering by a solid metallic cylinder [K. Gipson and P. L. Marston, J. Acoust. Soc. Am. 106, 1673 (1999)]. Free field observations of the backscattering by solid aluminum cylinders confirm that both meridional and helical rays contribute to the backscattering for different angles of cylinder tilt [K. Baik, Ph.D. thesis, WSU (2008)]. Those studies show that a previously developed criteria, based on the length of the cylinder, limits the range of tilt angles for which helical rays can contribute [F. J. Blonigen and P. L. Marston, J. Acoust. Soc. Am. 112, 528 (2002)]. Experiments with solid aluminum cylinders placed near a flat reflecting boundary suggest that the presence of a boundary can increase the number of helical ray paths that contribute to the backscattering; this increase can lead to increased backscattering amplitudes and an increased range of angles for which coupling to helical rays can occur. To investigate this effect, solid aluminum cylinders were suspended below the free surface of a water tank. Targets were illuminated from below at grazing incidence, and the backscattering was measured as a function of target rotation. [Work supported by ONR.]

MONDAY AFTERNOON, 19 APRIL 2010 ESSEX A/B/C, 1:00 TO 5:45 P.M.

Session 1pAA

Architectural Acoustics: Required Accuracy of Absorption, Scattering, and Diffusion Coefficients

Peter D'Antonio, Chair
RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Absorption, scattering, and diffusion coefficients: Overview. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

Acoustical design and computer modeling rely on the availability, accuracy, and use of coefficients which characterize how boundary surfaces affect incident sound. At this time three coefficients have been developed to characterize how incident sound is modified by absorptive and scattering surfaces. The random incidence absorption coefficient, which is a measure of the proportion of incident sound that is absorbed, is measured according to ISO 354/ASTM C423. It is well known that coefficients measured according to this standard suffer from several systematic errors which will be reviewed. Measurement of the random incidence scattering coefficient, which is the proportion of incident energy scattered in non-specular directions and used in computer modeling programs, is experimentally determined by ISO 17497-1. The limitations of this standard and applications will be discussed. Measurement of the diffusion coefficient, which is a measure of the uniformity of scattered sound and used to evaluate potential surfaces and in optimization programs, is described in AES-4id-2001 and ISO 17497-2 (pending). As with all standards it too has limitations which will be discussed. Detailed information on all three coefficients is given in Cox, T. J., and DAntonio, P. (2009). *Acoustic Absorbers and Diffusers* (Taylor & Francis, London).

1:25

1pAA2. Absorption coefficient error propagation into room acoustics parameters. Michael Vorländer (Inst. of Tech. Acoust., RWTH Aachen Univ. D-52056 Aachen, Germany)

It is well known that random-incidence absorption coefficients suffer from measurement uncertainties. Whether or not these uncertainties affect the quality of results from room acoustic prediction models depends on the specific case. One might argue that the absorption of the most relevant material must be known accurately. This, however, could be the material with highest absorption or the material with the largest surface area. In this presentation, the true random-incidence absorption coefficient and measurement uncertainties related to the current ISO 354 are briefly discussed. Then, in order to get a better view on the consequences of those uncertainties, quantitative uncertainties of some predicted parameters such as reverberation time, strength, and clarity are calculated by using an error propagation. Effects of uncertainties of absorption coefficients distributed in various room configurations are presented and discussed.

1:45

1pAA3. The true absorption—Measurements and coefficients. Christian Nocke (Akustikbuero Oldenburg, Katharinenstr. 10, 26121 Oldenburg, Germany, nocke@akustikbuero-oldenburg.de)

Many measurement procedures for sound absorption rely on approximations or necessitate special restrictions. The measurements in the tube usually only cover small samples and rely on a plane wave at normal incidence of sound. Reverberation chamber measurements assume diffuse sound fields and other constraints. Furthermore sound absorbers might show local or non-local reaction, e.g., angle dependence of the acoustic surface impedance or not. The performance of sound absorbers, either in a room or outdoors, often involves neither plane sound waves nor diffuse sound fields. This is one motivation for the development of measurement procedures applicable where the sound absorber is actually used, namely, *in-situ* methods. The paper will present a short historical review on the development of measurement techniques of sound absorption or the acoustic surface impedance. Measurement examples will be shown for some of the methods presented and compared to each other.

2:05

1pAA4. Investigation of the just noticeable difference of scattering coefficient in auralized concert halls. Renzo Vitale and Michael Vorländer (Inst. of Tech. Acoust., RWTH Aachen Univ., Neustrasse 50, 52066 Aachen, Germany, rvi@akustik.rwth-aachen.de)

The use of scattering coefficient in room acoustics simulations has proved to be an essential factor for improving the validity of results, as it quantifies the amount of the diffuse sound field in enclosed spaces. Measurement methods have been described in the norm ISO 17497. Available data, whose lack is still one of the weak points, are fairly growing. Moreover, the relationship between the physical quantity and perceptual aspects still remains unclear. Aim of this study is a first investigation toward the comprehension of the influence of scattering coefficient on the sound as perceived from the audience. The target is to quantify the just noticeable differences for variable scattering coefficients associated with lateral wall diffusers of a simulated concert hall. Several musical stimuli convolved with binaural room impulse responses are used as source signals for listening tests. Results focusing on perceptual aspects will be presented and discussed.

2:25

1pAA5. Characterizing scattering coefficients numerically via the fast multipole accelerated boundary element method. Nail A. Gumerov and Ramani Duraiswami (Perceptual Interfaces and Reality Lab., UMIACS, Univ. of Maryland, College Park, MD 20742, gumerov@umiacs.umd.edu)

Various panels are used in acoustical installations to provide desired characteristics to spaces to be used for listening. These panels have complex shapes with geometrical features that have sizes corresponding to wavelengths of sounds of interest. The complex interaction of acoustical waves with these shapes is what gives these surfaces their desirable properties. Experimental characterization of the acoustical properties of these surfaces under random and specular incidence is relatively time consuming. An alternate procedure is to numerically simulate the scattering behavior, and then computing the coefficients of interest from the simulation. A significant obstacle to such computations is the time taken for simulation. Fast multipole acceleration of boundary element methods [Gumerov & Duraiswami, J. Acoust. Soc. Am. **125** (2009)] is a promising approach to speeding up computations. We report on the application of this method to the computation of various scattering coefficients.

2:45

1pAA6. Evolution and uses of the diffusion coefficient: 1994–2010. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

A measurement method to determine the diffusion coefficient was first proposed at the Sabine Centennial Symposium in 1994. The method involved a boundary measurement technique which was carried out in a reflection free zone to determine the scattered impulse responses at 5 deg resolution from which the uniformity of third octave polar responses could be determined. In 2001, an information document (AES-4id-2001) was published describing this technique and the determination of the diffusion coefficient from the autocorrelation of the third-octave polar responses. This technique is being enshrined as ISO 17497-2, which will complement the scattering coefficient standard ISO 17497-1. This presentation will review the evolution of the diffusion coefficient, describe the current measurement and characterization technique, and detail its limitations and accuracy. Two examples of its use will be given. The first example will illustrate how the directional and random incidence diffusion coefficients can be used to evaluate the bandwidth and scattering uniformity of several potential one-dimensional and two-dimensional diffusing surfaces. The second example will demonstrate how the coefficient can be used as a metric to create new surfaces, using a shape optimization computer program, which combines boundary element and multi-dimensional minimization techniques.

3:05—3:20 Break

3:20

1pAA7. Comparison of measurement and simulation of the scattered pressure distribution from partly absorbing surfaces. Konstantinos Dadiotis, Jamie A. S. Angus, and Trevor J. Cox (Acoust. Res. Ctr., Univ. of Salford, Greater Manchester M37JT, United Kingdom)

When diffusion or scattering from a surface if discussed usually it refers to the distribution of pressure reflected from a rigid surface while the amount of reflected energy is not discussed. In this paper the scattered pressure distribution from surfaces that consist of absorbing and reflecting elements has been measured following the ISO 17497-2. The requirements of the standard were not perfectly

met as the signal to noise ratio was lower than required due to the absorbing nature of the surface and high-low-frequency noise in the measuring system. The measured results were then compared with simulations using boundary element modeling. The comparison showed good agreement. Finally the issue of the reflected energy was discussed.

3:40

1pAA8. Design of sound diffusion in concert halls using scale models. Jin Yong Jeon and Yong Hee Kim (Dept. of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

Effects of surface diffusion from lateral walls on concert hall acoustics were investigated to determine diffuser profiles. The design factors of diffusers in concert halls are geometrical shape, structural height, and surface coverage. Diffusion and scattering coefficients of the diffusers were measured in a 1:10 scale model reverberation chamber. Then, it was applied to scale model concert halls by design factors of diffusers. Acoustical parameters such as RT, EDT, C80, and sound pressure level (SPL) were calculated from the impulse responses. In addition, the numbers and amplitudes of the reflection peaks within −20 dB after the direct sound were calculated to evaluate the diffuseness in the halls. As a result, diffusers commonly yielded decrease in RT and SPL but contributed to increase in uniformity of RT and SPL. It was also found that the diffusive lateral surfaces close to the stage area are more effective in increasing both Np and ES. Appropriate design directions of diffuser are discussed from the measured results.

Contributed Papers

4:00

1pAA9. Characterizing rooms regarding reverberation time prediction and the sensitivity to absorption and scattering coefficient accuracy. B.-I. Dalenback (CATT, Mariagatan 16A, SE-41471 Gothenburg, Sweden, bid@catt.se) and Sarah Brown (RPG Diffusor Systems, Inc., Upper Marlboro, MD 20774)

Prediction of reverberation time (T30) using ray-tracing and similar algorithms naturally requires realistic input data, but the degree of data accuracy required varies with the room properties. In some types of rooms, accurate coefficients are crucial while in other types it is less crucial and the prediction results are less sensitive. This paper will attempt to classify rooms, according to their sensitivity to the input data due to their geometrical mixing (or diffusing) properties and the absorption, distribution. At the one end are found non-mixing geometries with uneven absorption, making a good estimate of the scattering coefficients crucial, at the other end are found mixing geometries with a mildly varying absorption distribution where prediction is safer and where estimated, or readily available, absorption coefficients can be used and where the choice of scattering coefficients has little impact. Due to the natural frequency dependence of both absorption and scattering coefficients the classification may also differ between octave-band predictions in the same room. Predictions are also difficult in rooms where the decay is not exponential. However, measures, e.g., STI and C80, may still be well predicted since they do not rely on fitting a straight line to a non-straight logarithmic decay.

4:15

1pAA10. Improving the accuracy of sound absorption measurements. Charles Moritz, Jennifer Shaw, and Armando Carrera (Blachford Inc., 1445 Powis Rd., West Chicago, IL 60185, cmoritz@blachfordinc.com)

Sound absorption test methods, such as ASTM C423, provide criteria for test chamber qualification. Measurements required during the ASTM C423 qualification process examine the variation in decay rates with microphone and test specimen locations. In addition, it is recommended to measure the change in sound absorption coefficients of a reference specimen with various amounts of diffusion in the room. However, there is little published information regarding systematic methods for determining diffuser, microphone, or specimen locations; therefore, qualification or improvement to the reverberation room depends on the acoustician's best judgment. As part of a continual improvement process to reduce testing variation, a mapping technique was utilized to visualize the three dimensional sound field in the room and determine changes necessary to improve the diffusion and repeatability of sound absorption measurements.

4:30

1pAA11. Quantifying the diffuseness of a sound field using a combined pressure and particle velocity sensor. Jonathan M. Botts, Philip W. Robinson (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY), Hans-Elias de Bree (Microflown Technologies and HAN Univ., Arnhem, The Netherlands), and Ning Xiang (Rensselaer Polytechnic Inst., Troy, NY 12180)

In many areas of acoustics, the degree to which a sound field is mixed, diffuse, or spatially uniform is important in understanding the system. In the case of performance spaces, it influences the audience's perception of the performance. More generally, it may have implications on the assessment of measurements taken in a space since often the evaluation of room-acoustic measurements relies on diffuse field assumptions. Several techniques have been proposed to quantify the diffuseness in a space, and one of the more recent methods will be investigated in this paper. Signals measured with a combined pressure and particle velocity sensor are used to calculate energy density and average intensity. The degree of diffuseness is then associated with the relationship between these two quantities. This paper discusses aspects of the relevant theory and practical implementation.

4:45

1pAA12. Simulating diffuseness with geometry variations of room surfaces. Sentagi S. Utami and Mojtaba Navvab (Dept. of Architecture, Univ. of Michigan, 2000 Bonisteel Boulevard, Ann Arbor, MI 48109, sentagi@umich.edu)

Several methods have been proposed to quantify the sound-field diffuseness utilizing measurements, room-acoustic modeling, and subjective assessment. Methods for measurements require sophisticated equipment with special consideration limited on quantifying uniformity of diffuseness in reverberant chambers. This study provides a systematic technique in simulating the sound-field diffuseness. The room simulated is typical rooms requiring need to control problems of echoes where the sound energy is often need to be conserved as well. Theoretical consideration on the sound source characteristics, source to receiver distances, geometrical size of the room surfaces, and material characteristics is taken into account and being altered in the simulation. Detail of the geometry variations in room surfaces is the major concern. Typical parameters in room acoustics are utilized to quantify the sound-field diffuseness. The results have shown that the geometry variations are correlated with the following energy properties: (1) reverberation times, (2) energies of the direct sound and reflections, (3) ratio of early-to-reverberant energies, and (4) the ratio of intensities of sound waves impinging on surfaces and reflecting back with its directional distributions.

5:00

1pAA13. On the accuracy of the subtraction method for *in-situ* reflection and diffusion coefficient measurements. Philip Robinson and Ning Xiang (Graduate Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

The subtraction method is a technique critical to measuring *in-situ* reflection/absorption coefficients as well as diffusion coefficients. In order to isolate a reflection, a reference measurement including only direct sound is subtracted from one with direct sound and the reflection of interest. However, taking two measurements with exactly the same direct sound component is complicated by environmental conditions, such as changes in temperature, air movement, and exact microphone positioning. These variations

can cause small time differences between the reference and the measurement which prevent complete subtraction of the direct sound; the residual direct sound then pollutes analysis of the isolated reflection. This work evaluates several methods to compensate for the differences to achieve minimal interference from the residual direct sound. A metric for gauging the success of the subtraction is proposed and the influence of the subtraction effectiveness on variations in the measured diffusion coefficient will also be discussed.

5:15

1pAA14. Inverse boundary and finite element formulations for the determination of high-frequency specific impedance. Alexander Bockman and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, xiangn@rpi.edu)

Material databases for consulting acousticians require accurate determination of specific impedance. Such data are typically gathered from the composition of data sets obtained from impedance tubes. In such tubes, the restriction of frequencies under test to those that satisfy a plane wave assumption for the tube geometry in question allows for the application of a one-dimensional transfer function analysis. Such tubes rarely achieve accuracy at the upper and lower ends of their operating spectra. Thus, the compositing of detailed data sets may produce significant error. For researchers of wave based methods of acoustic wave propagation, this represents a severe limitation. However, these data sets may be re-purposed for an extension of frequencies under test by the application of an inverse problem formulation of the boundary element method. The method accurately predicts data obtained within a larger impedance tube and further well extrapolates to data observed in a smaller impedance tube. Moreover, the method provides confidence in the consistency of specific impedance determination in composite data sets.

5:30

1pAA15. Higher-order impedance boundary conditions for finite difference solutions to the wave equation. Alexander Bockman, Jonathan Botts, and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

Higher-order finite difference methods for the solution of the wave equation are well established. The interchangeability of temporal and spatial derivatives allows for exact representations of the differential equation in difference forms to the order of accuracy desired. Realistic impedance conditions, however, germane to problems of an architectural acoustics origin, are typically given as ratios of pressure and velocity in the frequency domain. This work develops fourth order-accurate time domain boundary conditions for the types of materials typically encountered in the architectural acoustics field. Accuracy and agreement of numerical simulations to scale model measurements are compared and discussed, as is computational cost relative to lower-order methods.

MONDAY AFTERNOON, 19 APRIL 2010 WATERVIEW B/C, 1:00 TO 4:35 P.M.

Session 1pAB

Animal Bioacoustics: Effects of Anthropogenic Noise on Aquatic Animals II: Biological Damage

Carl R. Schilt, Chair
Bigleaf Science Services, P.O. Box 225, North Bonneville, WA 98639

Invited Papers

1:00

1pAB1. Preliminary analysis of effects of pile driving sounds on fish. Arthur N. Popper, Brandon M. Casper, Jiakun Song, David Sanderson-Kilchenstein (Dept. of Biology, Univ. of Maryland, College Park, MD 20742), Thomas Carlson (Pacific Northwest Natl. Labs., Battelle, Portland, OR 97204), and Michele B. Halvorsen (Univ. of Maryland, College Park, MD 20742)

There is continuing concern about the potential effects on fishes of the sounds produced during pile driving activities. Data in the literature are unclear as to physiological and behavioral effects. The earlier studies also generally did not include either experimental controls or control of signal parameters. We have developed a unique laboratory device that permits accurate simulation of impulsive pile driving signals in controlled experiments. Juvenile Chinook salmon were exposed to different combinations of sound levels and number of strikes to achieve various cumulative sound exposure levels. Controls received identical treatment as experimental animals, but without sound. The fish were evaluated by necropsy for assessment of barotrauma. Necropsies were conducted on exposed and control fish with experimenter not knowing the stimulus parameters. Preliminary analysis of results show that the extent of barotrauma rises with increased cumulative sound exposure level (SEL). Future experiments will examine effects of lower SELs as well as long-term effects that may arise at various time periods post-exposure. Ultimately, other species will be examined in order to develop broader criteria for exposure to pile driving sounds. (Work supported by NCHRP, CALTRANS, and MMS.)

1:20

1pAB2. Anthropogenic noise: Is this an issue for elasmobranch fishes? Brandon M. Casper and Arthur N. Popper (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, bcasper@umd.edu)

Underwater anthropogenic noise has developed into a major national and international issue since it is becoming more apparent that human activities can have an adverse effect on the survival of fish stocks. It is unclear, however, whether human-generated noise in the aquatic environment could have an effect on the health and behavior of elasmobranch fishes (sharks, skates, and rays), a group already being threatened worldwide due to overfishing. Many sources of anthropogenic noise produce sounds in the frequency range detectable

by these fishes; however, elasmobranchs are not considered to have sensitive hearing and therefore we do not know if they could be affected by these sounds. A review will be presented of what is known and unknown in regard to elasmobranch hearing abilities as well as behavioral studies in which sharks were exposed to or attracted by auditory stimuli. With these in mind, the goal will be to address whether anthropogenic environmental noise could be having deleterious effects on elasmobranch fishes.

Contributed Paper

1:40

1pAB3. Natural modes of vibration in the auditory system of a teleost. Carl R. Schilt (Bigleaf Sci. Services, P.O. Box 225, North Bonneville, WA 98639, schilt@saw.net), Ted W. Cranford (QMC Consulting, Inc., San Diego, CA), Petr Krysl (UCSD, San Diego, CA), Robert E. Shadwick (UBC, Vancouver, BC, CA), and Anthony D. Hawkins (Loughine, Ltd., Aberdeen, Scotland)

Fishes are diverse in hearing systems and anatomy. Doing actual bioacoustic experiments with even one species, candidate sound, and acoustic environment is difficult to conduct and to generalize. Our modeling approach, if properly carried out and validated, will enable us to do virtual bioacoustic experiments that will enhance knowledge of biophysical phenomena with economy of time and resources. We present preliminary results of a project that extends to fishes our finite-element modeling methods, which have been used to investigate marine mammal bioacoustics. These methods involve obtaining appropriate specimens, scanning them with x-ray CT, measuring elasticity of tissues, and constructing a finite-element model. Then we simulate the interactions between selected anthropogenic sounds and the model of the fish. Here we introduce vibrational analysis with fishes, simulating the resonant frequencies of the otoliths and surrounding structures in a large sciaenid fish (the white seabass, *Atractoscion nobilis*). Our preliminary analysis indicates the potential for complex relative motions among otoliths and, by implication, their sensory maculae.

Invited Papers

1:55

1pAB4. Metric for biological assessment of injury by percussive and decompressive exposure to impulsive sound. Thomas J. Carlson (Pacific Northwest Natl. Lab., 620 SW 5th Ave., Ste. 810, Portland, OR 97204-1423, thomas.carlson@pnl.gov)

Field and laboratory assessment of barotrauma to fish caused by percussion and/or decompression is complicated by the occurrence of injuries that do not result in immediate mortality. Accurate estimation of the consequence of exposure to impulsive sound must consider injuries that result in mortality delayed beyond typical post-exposure holding periods of exposed animals or that debilitate and lead to indirect mortality. Research of the response of fish to impulsive decompression over the past several years has resulted in development of a metric, mortal injury, that appears to more completely describe the biological cost to juvenile fish of exposure to rapid decompression. This measurement method has been extended to assessment of the response of juvenile fish to exposure to impulsive sound caused by blasting. The implications for assessment of take of listed species by blasting is evaluated by comparing mortality assessed over a 48-h holding period following exposure with that by necropsy for determination of mortal injury immediately following exposure.

2:15

1pAB5. Auditory hair cell regeneration in teleost fishes: A review. Michael E. Smith (Dept. of Biology, Western Kentucky Univ., 1906 College Heights Blvd., Bowling Green, KY 42101, michael.smith1@wku.edu)

Exposure to high levels of sound can cause damage to the auditory hair cells of many organisms, including fishes, resulting in hearing deficits. While such exposure can lead to permanent hearing loss in mammals, other phylogenetically older nonmammalian vertebrates, such as birds, reptiles, amphibians, and fishes, can spontaneously regenerate hair cells in both the vestibular and auditory portions of the inner ear following trauma. The general mechanism of hair cell regeneration is currently being examined in a number of different taxa, but data are limited on hair cell regeneration in the inner ear of fishes. This presentation will summarize the few studies that have examined auditory hair cell regeneration in teleost fishes and compare the results reported for fishes to those reported for other species. While understanding the process of hair cell regeneration in fishes has biomedical relevance, this knowledge could also benefit marine managers attempting to mitigate the effects of anthropogenic noise on fishes. [Work supported by NIH P20 RR16481.]

2:35

1pAB6. Noise-induced hearing loss: From animals to humans. Colleen Le Prell (Dept. of Communicative Disord., Univ. of Florida, Box 100174, Gainesville, FL 32610, colleeng@phhp.ufl.edu)

Exposure to noise can result in neural swelling, mechanical damage, and/or a toxic accumulation of free radicals. The type and extent of the pathology, the amount of hearing loss, and the permanence of the hearing loss all vary with the level and duration of the noise insult. In addition, noise exposures in which sound level rapidly changes over time are more harmful than relatively constant level exposures. These relationships have been carefully worked out in animal models, and key findings will be reviewed. Evaluating the effects of noise on the human inner ear is more challenging. Exposures in the work place and other real-world environments can be challenging to control, and in those studies evaluating long-term effects of noise, the subjects are exposed to variable non-study noise sources. Exposures in the laboratory that induce temporary changes in human hearing carry their own challenges. In this lecture, we will discuss various human models for studies on the effects of noise, and we will present new data on temporary hearing changes in human subjects that listen to music with digital audio players. Several metrics used to measure the effects of noise will be discussed.

2:55—3:15 Break

3:15

1pAB7. Effects of mid-frequency sonar on fish. Michele B. Halvorsen (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, mb_halvorsen@yahoo.com), David A. Zeddies (Marine Acoust. Inc., Arlington, VA 22203), William Ellison (Marine Acoust. Inc., Middletown, RI 02482), Jiakun Song (Univ. of Maryland, College Park, MD 20742), David R. Chicoine (New York Chiropractic College, Seneca Falls, NY 13148), and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

There is significant concern that high-intensity sounds can impact fish physiology and behavior. Important sources of such sounds are commercial and military sonars. A recent publication reports that exposure to low-frequency active (LFA) sonar might result in hearing loss in some fish species, but did not damage any tissues [Popper *et al.* (**2007**). J. Acoust. Soc. Am. **122**, 623–635]. We have extended our studies to include U.S. Navy mid-frequency sonar. Fish were exposed to a 3-s-long signal that consisted of a 2-s 2.8–3 .8-kHz frequency sweep, immediately followed by a 1-s 3.3-kHz tone). The stimulus was repeated five times with a 25-s interval. Fish were subsequently tested for hearing sensitivity and examined both grossly and microscopically for tissue damage. Some temporary hearing loss was found in catfish, species known to hear sounds above 1000 Hz, whereas there was no effect in fish which do not hear above about 1 kHz. There was no gross damage to any tissue, and microscopic examination showed no effect on any tissues. [Work supported by Office of the U.S. Chief of Naval Operations.]

Contributed Papers

3:35

1pAB8. Noise dose for aquatic animals: Preliminary estimates for two seismic surveys. Nils Olav Handegard (Inst. of Marine Res., NO-5817 Bergen, Norway, nilsolav@imr.no)

In humans, a daily noise dose is defined and used to regulate noise exposure in the working environment. This is an integrated measure of sound intensity over the working day, typically 8 h in duration. When investigating behavioral effects on aquatic animals exposed to anthropogenic noise, a comparable measure would be valuable, especially when transient but powerful underwater noise sources such as seismic air-guns and mid- and low-frequency sonars are being used or are being considered for use. Here, an integrated noise dose measure is defined. Preliminary estimates of this noise-dose measure are given for two different seismic surveys.

3:50

1pAB9. Assessing and regulating the impact of sound and vibration upon fish. Anthony Hawkins (Loughine Ltd., Kincraig, Blairs, Aberdeen AB12 5YT, United Kingdom, a.hawkins@btconnect.com)

The generation of very high-sound levels in water has the potential to change the behavior of fish and in extreme cases cause injury. Such sounds may be generated by pile driving, seismic exploration, low-frequency sonar equipment, ships, and underwater dredging and drilling activities. Some land-based activities like drilling, excavation work, and vibroseis may also generate sound and vibration which subsequently enters adjacent water bodies like lakes and rivers. It has become commonplace for regulatory agencies to require the impact of these sound-making activities to be formally assessed for their impact upon aquatic animals, including fish. The procedures for carrying out these assessments vary and, in many instances, are poorly performed or subject to uncertainty. The uncertainties which come into play when impacts upon fish are assessed include lack of knowledge on the sensitivity and response of fish to sounds, the physical conditions surrounding sound generation and propagation, and the metrics to be applied to the stimulus received by the fish. This paper will consider those sound exposure criteria applied to fish for regulatory purposes and how they might be refined.

4:05

1pAB10. A model for predicting effects of sound on auditory sensory cells in fishes. Mardi C. Hastings (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, mardi.hastings@gatech.edu)

Relationships among peripheral auditory mechanics, characteristics of sound exposure, and effects on auditory tissues in fishes are not understood. Several studies in the literature have reported varying degrees of damage to auditory hair cells in different species exposed to different types of sounds; however, there appears to be no correlation between sound exposure level and observed damage. In effort to understand mechanisms underlying hearing loss and hair cell damage and to predict their occurrence, a mathematical model of the peripheral auditory system in fishes was developed to study the dynamic response of the ear to incident sound in five different species. This model determines relative motion between the sensory epithelium and otolith in the inner ear, which causes bending of the apical ciliary bundles of auditory hair cells. Excessive relative motion predicted by the model was correlated with hair cell damage and found to be in good agreement with results of studies reported in the literature. Swim bladder geometry, otolith size, anatomical connections between swim bladder and inner ear, and spectral characteristics and amplitude of the received sound were found to affect the degree of hair cell damage. [Work supported by the International Association of Oil and Gas Producers.]

4:20

1pAB11. Assessment of the acoustic effects on marine animals by an offshore wind farm. James H. Miller, Gopu R. Potty, Kathleen Vigness Raposa, David Casagrande, Lisa A. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, miller@uri.edu), Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105), and Peter M. Scheifele (Univ. of Cincinnati, Cincinnati, OH 45267)

As part of the planning for an offshore wind farm in Rhode Island coastal waters, an assessment of the potential acoustic effects on the ecosystem is being conducted. The developer has proposed to initially deploy eight 3.6-MW wind turbines within 3 nm of Block Island. Two passive aquatic listener (PAL) systems were deployed south of Block Island from October 6 to November 11, 2008. Using data from the PALs, ambient noise histograms were computed for this pre-construction phase. The largest sources of noise in the area at low frequencies were found to be from shipping, wind, rain, and biological sources. In addition, transmission loss measurements were also made in the region to calibrate a geoacoustic model. Measurements of airborne noise from a 1.5-MW land-based wind turbine already in operation in Rhode Island were made in 1/3-octave bands and near the proposed wind-farm site. A preliminary assessment of the effects of the offshore wind farm on marine animals at these sites will be presented. A plan for monitoring the noise field and potential biological effects during construction and operation of the windfarm is presented. [Funding provided by the RI Office of Energy Resources.]

Session 1pAO

Acoustical Oceanography, Animal Bioacoustics, and Underwater Acoustics: Acoustics in Polar Environments II

Jennifer L. Miksis-Olds, Cochair
Pennsylvania State Univ., Applied Research Labs., P.O. Box 30, State College, PA 16804

David K. Mellinger, Cochair
Oregon State Univ., 2030 S.E. Marine Science Dr., Newport, OR 97365

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Invited Papers

1:00

1pAO1. Marine bioacoustics in a changing Arctic. Robert C. Gisiner (Marine Mammal Commission, Rm. 700, 4340 East-West Hwy., Bethesda, MD 20814, bgisiner@mmc.gov)

The retreat of summer ice in the Arctic during the past 5 years has been accompanied by a surge of alarm about iconic ice-associated marine mammal species: polar bears, walrus, ringed seals, and others. Changes for subarctic species of mammals, fishes, and planktonic organisms have not received as much attention, but will also need to be monitored. The paucity of data about Arctic marine mammals is now more than ever a serious impediment to good decision making. One of the greatest sources of hope is the tremendous advance in marine bioacoustic sensing technology that has occurred within the past decade. Passive and active acoustic technologies offer new means of obtaining biological data in a summer Arctic Ocean where ice is less available as a platform for animals and scientists, and in a winter Arctic Ocean that will continue to be one of the most difficult research environments in the world.

1:20

1pAO2. Including passive acoustic capability in Arctic ocean observing systems. Sue E. Moore (NOAA/Fisheries S&T, 7600 Sand Point Way NE, Seattle, WA 98115, sue.moore@noaa.gov), Kathleen M. Stafford (Univ. of Washington, Seattle, WA 98105), Catherine L. Berchok (NOAA, Seattle, WA 98115), Humfrey Melling (Inst. of Ocean Sci. DFO, Sidney, BC V8L 4B2, Canada), and Oystein Wiig (Univ. Oslo, Blindern, N-0318 Oslo, Norway)

Over the past decade, long-term deployments of passive acoustic recorders have provided a new baseline on the seasonal occurrence of large whale species in remote regions of the world ocean. In the Arctic, passive acoustic sampling has identified both whale calls and sounds from anthropogenic sources (ships and seismic profiling), activities that are expected to increase with diminished sea ice cover. In 2008, NOAA capitalized on an opportunity to join on-going IPY projects by inclusion of recorders at three High Arctic mooring sites: one on the Chukchi Plateau and two on either side of Fram Strait. The recorders (AURAL-M2) provided a year of sub-sampled (9 min on/11 min off) recordings at 0.1 Hz to 4 kHz (8192 sampling rate), which encompasses the bandwidth of whale and ice seal calls. Data from the recorders were complemented by a suite of standard oceanographic measures from other instruments on the mooring line. Provisional results show novel occurrence of both marine mammal and anthropogenic signals in the High Arctic. To realize the vision of a Global Ocean Acoustic Observing Network [Dushaw *et al.* (2009)], passive acoustic technology must become a standard sampling component, especially in the Arctic during this time of rapid climate change.

1:40

1pAO3. Effects of seismic exploration activities on bowhead whale call distribution in the Alaskan Beaufort Sea. Susanna B. Blackwell (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117, susanna@greeneridge.com), Christopher S. Nations, Trent L. McDonald (WEST, Inc., Cheyenne, WY 82001), Aaron Thode (Scripps Inst. of Oceanogr., La Jolla, CA 92093), Katherine H. Kim, Charles R. Greene, Jr. (Greeneridge Sci., Inc., Goleta, CA 93117), and A. Michael Macrander (Shell Exploration & Production Co., Anchorage, AK 99503)

Airgun use during seismic exploration activities is known to cause a decrease in call detection rates (used here as a proxy for calling rates) in migrating bowhead whales. However, the received airgun sound levels at the whales, at which behavioral changes occur, are not known. To address this question 40 directional autonomous seafloor acoustic recorders (DASARs) were deployed in autumn 2008 at five sites along the migration corridor of bowhead whales during ongoing marine seismic exploration activities. Over 440 000 whale calls were detected and localized by triangulation. Concurrently, over 100 000 airgun pulses were detected and analyzed. The study area was divided into a hexagonal grid (hexagon width 1.75 km). The received level (SPL and cumulative SEL) of airgun sounds was modeled for each hexagon in the study area and each 15-min period over the entire season, using information on the timing and location of airgun activities, the size of the airgun array being used, the depth of the DASARs recording the airgun sounds, and other covariates. Logistic regression was then used to model occurrence of whale calls as a function of received level, distance offshore, and other covariates. [Work supported by Shell Exploration and Production Company.]

2:00

1pAO4. Are they really not there? Using passive acoustics to overcome false absences in the study of vocal species that are rare, secretive, or distributed at low densities. Tracey L. Rogers, Carolyn J. Hogg, Michaela B. Ciaglia (Evol. and Ecol. Res. Ctr., School of BEES, Univ. of NSW, Sydney 2052, Australia tracey.rogers@unsw.edu.au), and Douglas H. Cato (Defence Sci. and Technol. Organisation, New South Wales, Australia)

Estimating abundance and spatial use behavior can be challenging for marine species that are rarely sighted. This situation is exacerbated in the polar regions due to the peculiar logistical difficulties of working in the pack ice, which makes survey effort enormously expensive. Presented is a simple approach for modeling sounds per animal over a unit time as a relative index for species where there is information on the production of vocalizations (acoustic behavior including seasonal calling patterns, diurnal calling patterns, inter-individual stereotypy, inter-sexual stereotypy, audience effect, and predictable calling rate over a unit of time) and the detection range of those vocalizations (survey distance—theoretical estimates calculated with call intensities). We focus on an Antarctic pack ice seal, the leopard seal, and Hydrurga leptonyx, as estimating abundance from survey effort faces challenges. Our case study shows that with the advent of more sophisticated marine engineering coupled with effort focused on the pertinent parameters of acoustic behavioral ecology we can open up the scope to study the behavior of rare, secretive, or low-density species across a range of site occupancy studies, density, and habitat use at a time when we need to develop cost efficient tools to glean an understanding of what is happening in these most vulnerable locations.

2:20

1pAO5. From psychophysics to management of noise-disturbance in a large, Arctic carnivore, the polar bear. Ann E. Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.com) and Megan A. Owen (San Diego Zoo Inst. for Conservation Res., San Diego, CA 92112-0551)

About 50% of maternal dens excavated by pregnant female polar bears on Alaska's North Slope occur on land or land-fast ice. Management agencies and the public have raised concerns that noise from human activities could adversely affect denning female polar bears, causing den abandonment, interference with communication, altered habitat use, or behavioral and physiological stress. Although US Fish and Wildlife Service uses approach limits to prevent disturbance of denning females that are assumed to be conservative, no science-based standards have been developed and uncontrolled human activities are still possible. The potential for effects will be reviewed in light of published evidence and measurements of polar bear auditory thresholds [Bowles *et al.* (2008). J. Acoust. Soc. Am. **123**, 3509] to show how an understanding of species-specific psychophysical characteristics is needed to assess the impact of noise disturbance from industrial activities. Scientific challenges to quantifying effects include identifying efficient noise metrics and practical methods for measuring adverse responses. The negative ramifications of disturbance may be intensified due to anticipated reductions in polar bear physical condition as their sea ice habitat declines.

2:40

1pAO6. Air-borne noise in the Arctic and implications for polar bears. Alexander O MacGillivray (JASCO Appl. Sci., 2101-4464 Markham St., Victoria BC, V8Z 7X8 Canada) and Christine Erbe (JASCO Appl. Sci., Bellbowrie, Queensland 4070, Australia)

An increasing number of Arctic animal populations are listed as threatened. The list includes ungulates, terrestrial carnivores, marine mammals, and birds. Potential threats are habitat degradation and loss, pollution, prey depletion, harvesting pressure (as a result of human population growth), climate change, and noise. Growing exploration and industrialization of the Arctic come with an increase in underwater and air-borne noise. Sources of noise relate to oil, gas and mineral exploration and production, hydroelectric schemes, construction (of ports, buildings, roads), tourism, and traffic (marine, land-based and air). Noise and vibration measurements of land-based industrial activities will be presented, including excavation equipment, various types of ground vehicles, and blasting. Sound propagation characteristics will be explained. Results of a study measuring the transmission of industrial noise into polar bear dens will be shown and the potential effects of noise on Arctic animals will be discussed.

3:00—3:30 Break

Contributed Papers

3:30

1pAO7. Acoustic propagation effects on airgun pulses and whale calls in the Beaufort Sea. Katherine H. Kim (Greeneridge Sci., Inc., 1411 Firestone Rd., Santa Barbara, CA 93117, khkim@greeneridge.com), Aaron Thode (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), and Charles R. Greene, Jr. (Greeneridge Sci., Inc., Santa Barbara, CA 93117)

In 2008, vocalizing bowhead whales (*Balaena mysticetus*) conducted their fall migration in the Beaufort Sea amidst short- and long-range seismic exploration activities. Whale calls and airgun pulses were recorded on five arrays of Directional Autonomous Seafloor Acoustic Recorders (DASARs) deployed along 280 km of Alaska's North Slope. One particular sequence of pulses was recorded from the *Louis S. St. Laurent*, a Canadian geophysical vessel conducting surveys in deep offshore waters in collaboration with the United States Geological Survey (USGS), thus providing a fix on the location of some distant airgun activity. The distant airgun signals display an interesting selection of long-range propagation effects, which will be discussed and analyzed in the context of range-dependent propagation modeling in a combined deep-water upward-refracting and shallow-water bottom-interacting environment. [Work supported by the U.S. Geological Survey.]

3:45

1pAO8. Ambient noise in the Chukchi Sea, July 07 Oct 2009. Bruce Martin, Julien Delarue, and David Hannay (JASCO Res., 1496 Lower Water St., Halifax, NS B3B 1R9, Canada, bruce@jasco.com)

Shell Exploration and Production Company and ConocoPhillips Alaska have commissioned a multi-year acoustic study of marine mammal vocalization activity and ambient sound levels in the Alaskan Chukchi Sea. A large acoustic dataset, collected between July 2007 and October 2009, has been analyzed and has provided a wealth of information about the seasonal soundscape of the Chukchi Sea. This paper presents the ambient noise results throughout the acquisition time period. The analysis investigates the contributions made to ambient levels by shipping, marine mammals, seismic activity, and ice. We will show how short term weather conditions can

strongly affect the noise levels measured on bottom-moored recorders. Certain mammal vocalizations such as bearded seal trills and bowhead moans can dominate ambient noise levels over large time periods.

4:00

1pAO9. Where do the Chukchi Sea fin whales come from? Looking for answers in the structure of songs recorded in the Bering Sea and Western North Pacific. Julien Delarue (JASCO Appl. Sci., 432-1496 Lower Water St., Halifax, NS B3J 1R9, Canada), Dave K. Mellinger (Oregon State Univ., Newport, OR 97365), Kathleen M. Stafford (Univ. of Washington, Seattle, WA 98105), and Catherine L. Berchok (NOAA/AFSC, Seattle, WA 98115)

Fin whales are common throughout the North Pacific and have recently been detected acoustically as far north as the northeastern Chukchi Sea. Non-acoustic evidence suggests that North Pacific fin whales are segregated into two populations wintering along the Asian and North American coast with at least some animals intermingling in the summer in the Bering Sea-Aleutian Islands area. Male fin whales produce regionally distinctive songs which are likely indicative of population structure. In this study we evaluated the relationship of fin whales recorded in the northeastern Chukchi (2007 and 2009) and southeastern Bering (2007–2008) seas by comparing the structure of their song. Additionally, we investigated whether fin whales detected in these areas could be part of an Asian population by comparing their songs to those recorded near the Emperor Seamounts in the western North Pacific (2007). The results will be discussed in light of the current knowledge on North Pacific fin whale population structure.

4:15

1pAO10. Factors influencing biodiversity and marine animal habitat use in the Bering Sea. Jennifer L. Miksis-Olds (Appl. Res. Lab., Penn State, P.O. Box 30, State College, PA 16804, jlm91@psu.edu) and Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105)

A combination of active and passive acoustic technology was integrated into NOAA oceanographic moorings at two locations on the eastern Bering Sea shelf. Physical, biological, and acoustical oceanographic instruments made synoptic measurements of marine mammal presence, zooplankton/fish concentrations, physical oceanographic processes, primary production, and sound levels from 2008–2009. Synergistic factors influencing biodiversity and marine animal habitat use were investigated. Observed patterns in the central Bering Sea indicate that ecosystem dynamics and the acoustic environment are strongly driven by sea ice. Marine mammal presence and diversity were tightly coupled to sea ice presence. Bowhead whales and ice seals dominated the acoustic soundscape during the winter and spring, whereas geophysical signals dominated during the fall and summer months. Zooplankton abundance was also tightly linked to ice coverage. Time series of acoustic backscattering strength during the winter months indicated that the zooplankton community rapidly responds to short period of open water at a level that equals or exceeds maximum volume backscatter values at other times of the year. Integrated data such as these will be vital in understanding the relationship between marine mammals, zooplankton, and the physical environment. [Work supported by ONR.]

4:30

1pAO11. Analysis of marine mammal detections from the Bering Sea relative to a 10 day retreat of sea ice in March 2009. Samuel L. Denes (Grad. Prog. in Acoust. and Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804-0030, sld980@psu.edu) and Jennifer L. Miksis-Olds (The Penn State Univ., State College, PA 16804-0030)

Patterns of marine mammal detections from an adaptive sub-sampling acoustic recorder deployed in the Bering Sea are related to a temporary retreat of sea ice. Ice coverage data from the NOAA Ice Desk showed a 10 day period in March 2009 during which no ice was detected over an NOAA oceanographic mooring equipped with the acoustic recorder. During periods of ice coverage, bearded seal and bowhead whale vocalizations dominated the soundscape and were the primary sources triggering a high-duty cycle sampling mode. Bearded seal vocalizations decreased immediately following the transition from ice cover to open water. The sub-sampling protocol limits the number of short duration sound clips recorded each day in order to conserve power and storage space during year deployments. Periods with ice coverage preceding and following the ice retreat exceeded the maximum daily limit within the first 2 h of the day. During the first day of open water, there were only two detections. Temporal variation in the number of bearded seal detections is highly coupled to sea ice characteristics, illustrating the importance of this environmental feature to the bearded seal mating system. [Work supported by ONR.]

4:45

1pAO12. Right whale distribution in the Bering Sea revisited: A fresh look. Jessica Crance, Catherine Berchok, Phillip Clapham (NOAA Natl. Marine Mammal Lab., 7600 Sand Point Way, Seattle, WA 98115, jessica.crance@noaa.gov), Marc Lammers, Whitlow W. L. Au (Hawaii Inst. of Marine Biology, Kaneohe, HI 96744), David K. Mellinger, and Sharon Nieukirk (Oregon State Univ., Newport, OR 97365)

The North Pacific right whale (*Eubalaena japonica*) is one of the most endangered baleen whales in the world, with current abundance estimates of fewer than 40 individuals. As such, it has been the focus of intensive population monitoring studies. Long-term passive acoustic monitoring is a useful tool for assessing distribution, habitat use, and population status of marine mammals. For Bering Sea right whales, past analyses of distribution were based on their stereotyped upsweep call. Results from sonobuoy recordings made during the 2008 and 2009 southeastern Bering Sea right whale cruises, however, found gunshot calls to be much more common. To assess whether inclusion of gunshot call detections will change our understanding of the spatial and temporal distribution of North Pacific right whales, we examined four years (2006–2009) of long-term recordings made by the National Marine Mammal Laboratory and the Pacific Marine Environmental Laboratory at six different locations on the Bering Sea shelf. A comparison of the results using upsweeps, gunshots, and a combination of both will be presented. [Work supported by an Inter-Agency Agreement from the Minerals Management Service.]

5:00

1pAO13. Results and insights from operational acoustic monitoring networks in the Chukchi Sea, Summer-Fall 2006 and 2008. Russell A. Charif, Ann Warde, Dimitri Ponirakis, Anita Murray, Charles Muirhead, Michael Pitzrick, Christopher W. Clark (Bioacoustics Res. Program, Cornell Univ., 159 Sapsucker Woods Rd., Ithaca, NY 148504), Sheyna Wisdom (URS Corp., Anchorage, AK 99503), and Caryn Rea (Conoco Phillips Alaska, Inc., Anchorage, AK 99501-3439)

Recordings for marine mammals were collected in the Chukchi Sea in the 2006 and 2008 summer-fall seasons in proximity to seismic exploration activities. 2006 data were from MARUs off Cape Lisburne, Pt. Lay, Wainwright, and Pt. Barrow, AK. 2008 data were from MARUs in the Burger and Klondike areas. In 2006, belugas, but no bowheads, were rarely detected off Cape Lisburne in July–September. Bowheads, but no belugas, were detected off Wainwright, Lay, and Barrow in September–October, with bowheads very common off Barrow. Seismic airgun sounds were common on distant MARUs off Cape Lisburne, Lay, and Wainwright. There were very few seismic events detected off Barrow and those few were all close to 0-dB SNR. In 2008, bowheads were more common at Burger than Klondike and occurred mostly during the latter part of the September–October period; belugas were not detected at Burger and only on 6 days early in the period at Klondike; walrus were more common at Burger than at Klondike. In both areas bowhead densities were low-moderate. Seismic sparker sounds were detected on 7 and 28 days at Burger and Klondike, respectively, with higher received levels at Klondike.

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Therapeutic Ultrasound Applications

Diane Dalecki, Cochair
Univ. of Rochester, Dept. of Biomedical Engineering, 310 Goergen Hall, Rochester, NY 14627

James McLaughlan, Cochair
Boston Univ., Dept. of Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Chair's Introduction—12:55

Contributed Papers

1:00

1pBB1. A two-scale model to describe cavitation dynamics in viscoelastic media. Stéphane Labouret, Jamie R. T. Collin, and Constantin C. Coussios (Biomedical Ultrason. & Biotherapy Lab. (BUBL), Inst. of Biomedical Eng., Dept. of Eng. Sci., Univ. of Oxford, Headington, Oxford OX3 7DQ, United Kingdom, stephane.labouret@eng.ox.ac.uk)

In spite of several decades of research on cavitation inception and onset in biological media, the exact environment experienced by a cavitating bubble at the micro-scale and in the MHz frequency range remains poorly understood. Comparison of model-based predictions of bubble behavior with optical observations in transparent media suggests that the mechanical property values of the medium could be up to two magnitudes different than the corresponding static values. Several viscoelastic media, such as tissue or tissue-mimicking materials, have been recently shown to exhibit a submicron structure which can be assumed equivalent to a pure liquid up to a characteristic length scale and essentially viscoelastic thereafter. The present work proposes an extension to the model recently proposed by Yang and Church [J. Acoust. Soc. Am. (2005)] that takes this dual scale into account. Model predictions are validated against experimental results obtained by inducing single-bubble inertial collapses in an agar-based tissue mimicking material using short pulses produced by a 1.1-MHz HIFU transducer; the resulting acoustic emissions are acquired using a calibrated passive cavitation detector, thus enabling direct comparison with the inertial cavitation threshold and radiated sound amplitude predicted by the single-scale and novel dual-scale models.

1:15

1pBB2. High-intensity focused ultrasound as a potential treatment modality for heterotopic ossification. Sandra L. Poliachik (Orthopaedics and Sports Medicine, Univ. of Washington, 325 Ninth Ave., Seattle, WA 98104, poliachi@u.washington.edu), Tatiana D. Khokhlova, and Michael R. Bailey (Univ. of Washington, Seattle, WA 98105)

Heterotopic ossification (HO) is the growth of bone in abnormal locations such as soft tissue. HO is a frequent secondary complication of common trauma, e.g., amputation or spinal cord injury. HO causes pain, motion restriction, and poor prosthetic fitting. Current HO treatments include nonsteroidal anti-inflammatory drugs, radiation, surgery, and shock-waves, but HO often returns. In this work, feasibility of HO treatment using high-intensity focused ultrasound (HIFU) is studied. HIFU may induce destruction of HO nodules by a combination of thermal ablation and mechanical erosion through bubble activity. In preliminary experiments bone sections from rodent proximal tibiae embedded in polyacrylamide gel were used as tissue phantoms; subsequent experiments were performed on HO induced in a mouse calf muscle. A 2-MHz focused transducer was used to generate pulses of shock waves with peak positive pressure of 67 MPa, duration of 10–100 ms, and 1%–10% duty cycle. High-resolution microCT imaging of the samples was performed prior to and following HIFU exposure. Both mechanical damage of the porous bone structure and thermal denaturation of the bone were observed. The data obtained suggest that HIFU may be an option for non-invasive treatment of HO. [Work supported by NSBRI _SMSTO1601 and NIH_DK007742].

1:30

1pBB3. Nonlinear distortion associated with high-intensity focused ultrasound creates apparent displacements in acoustic radiation force imaging. Andrew B. Draudt and Robin O. Cleveland (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02474)

An acoustic radiation force imaging (ARFI) system has been developed which employs a 1.1-MHz high-intensity focused ultrasound (HIFU) source to create the pushing beam (0.5–7-MPa peak pressure) and a focused 7.5-MHz transducer (mounted concentrically in a hole of the HIFU transducer) to create the imaging pulse. The motion of tissue is determined by recording multiple A-lines from the imaging transducer and using speckle tracking to estimate displacement. The HIFU was amplitude modulated to result in a time-varying radiation force and imaging pulses transmitted simultaneously in order to monitor tissue motion during the push phase. Measurements of the pressure waveforms in water indicated that nonlinear distortion of the HIFU could advance the arrival time of the imaging pulse if it was on a pressure peak and retard it if it was on a pressure trough, resulting in apparent displacements on the order of 5 μm. The apparent displacement varied with the phase at which the ARFI was transmitted and increased with the HIFU pressure amplitude. Experiments with ARFI in chicken breast indicated that the apparent displacement varied with depth which appeared to be related to differences in the diffraction patterns of the two sources. [Support NSF CenSSIS EEC-9986821.]

1:45

1pBB4. Rytov modeling for backscatter monitoring of high-intensity focused ultrasound therapy. Gavriel Speyer, Peter Kaczkowski, Andrew Brayman, and Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Accurate monitoring of high-intensity focused ultrasound (HIFU) therapy is critical for ensuring proper treatment. Changes in time of flight for diagnostic ultrasound (DU) backscatter have been recognized as a means to identify temperature changes between two DU frames, captured before and after HIFU therapy. Estimates of heating from precisely controlled HIFU experiments strongly indicate that DU measurements are formed by an interaction between the heated material and a diffuse DU beam, as opposed to the commonly assumed ray model for the DU beam. The existence of a diffuse beam can be explained using the Rytov method, where temperature induced changes in material properties affect the phase and amplitude of the probing signal, with cumulative effects produced over characteristic volumes, and not simply depths. In particular, the temperature effect for on-axis backscatter is shown to be a generalization of the models proposed by Simon (1998) and Moreno (1996), with the phase determined by a complex weighted average of temperature in cross sections perpendicular to the DU

beam axis. Results are illustrated using data obtained from gel phantoms, where the beam interaction is shown to provide therapy estimates consistent with the true therapy. [Work supported by NIH grant 5R01CA109557].

2:00

1pBB5. Tissue erosion using millisecond boiling in high-intensity focused ultrasound fields. Michael S. Canney (INSERM U556, 151 Cours Albert Thomas, Lyon Cedex 03, France), Tatiana D. Khokhlova, Yak-Nam Wang, Vera A. Khokhlova, Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

High-intensity focused ultrasound (HIFU) transducers can be operated at high-pressure amplitudes of greater than 60 MPa and low-duty cycles of 1% or less to induce controlled bubble activity that fractionates tissue. The goal of this work was to investigate fractionation not from mechanically induced cavitation but from thermally induced boiling created by HIFU shock waves. Experiments were performed using a 2-MHz HIFU source. The focus was placed in *ex vivo* bovine heart and liver samples. Cavitation and boiling were monitored during exposures using a high-voltage probe in parallel with the HIFU source and with an ultrasound imaging system. Various exposure protocols were performed in which the time-averaged intensity and total energy delivered were maintained constant. The types of lesions induced in tissue ranged from purely thermal to purely mechanical depending on the pulsing protocol used. A pulsing protocol in which the pulse length was on the order of the time to boil (of only several milliseconds) and duty cycle was low ($<1\%$) was found to be a highly repeatable method for inducing mechanical effects with little evidence of thermal damage, as confirmed by histology. [Work supported by NIH EB007643, NSBRI SMST01601, and RFBR 09-02-01530.]

2:15

1pBB6. Enhanced high inensity focused ultrasound-induced heating in a phantom. Cecille Pemberton Labuda (Dept. of Phys. Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, cpembert@olemiss.edu)

High intensity focused ultrasound (HIFU) has the capability to induce hemostasis through rapid delivery of energy in the HIFU focal region. HIFU effectively induces hemostasis in small blood vessels; however, as vessel size increases, loss of heat to the flow increases until HIFU treatment is only minimally effective. A new method to enhance the HIFU-induced heating effect was investigated. Single polymer fibers embedded in a heat-sensitive phantom are shown to increase the size of HIFU-induced lesions as compared to controls. Polymer meshes were then embedded in the phantom and the size of the HIFU-induced lesions produced were compared with control lesions.

2:30

1pBB7. Transcranial microbubble destruction by using re-focusing ultrasound. Po-Wen Cheng (Graduate Inst. of Biomedical Electrons and Bioinformatics, Natl. Taiwan Univ., Rm. 731, MD Bldg., No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106, f95921110@ntu.edu.tw) and Pai-Chi Li (Natl. Taiwan Univ., Taipei, Taiwan 106)

Administration of microbubbles (MBs) during transcranial Doppler monitoring further accelerates ultrasound-enhanced thrombolysis. However, low efficiency of transcranial ultrasound transmission resulting from phase aberrations through skull is still an obstacle. To alleviate this, re-focusing by phase aberration correction using echoes from microbubbles can be performed. Two re-focusing methods were designed for transcranial MB destruction. One is adjusting the emission phases derived from the channel-to-channel correlation of received signals. The other is simply adjusting the transmit focal depth. This is possible because the phase aberrations typically result in approximately parabolic delay errors. Focusing quality was compared using finite difference time domain simulations, and the mechanical index (MI) distributions were calculated as indices for likelihood of microbubble destruction. Using the correlation based methods, the beamwidth becomes 45.5% narrower than that by using the time reversal method. The MI can also be increased by 15.9% while keeping it under safety range. Adjusting the transducer position, on the other hand, can also enhance the focusing efficiency and improve efficiency of ultrasound-enhanced thrombolysis. It is also easy to implement using current clinical systems.

2:45

1pBB8. Noninvasive *in vivo* cavitation threshold detection during blood-brain barrier opening using focused ultrasound and the contrast agent and definity. Yao-Sheng Tung, Fotis Vlachos, Thomas Deffieux, and James Choi (Dept. of Biomedical Eng., Columbia Univ., Vanderbilt Clinic 12-232, 622 W. 168th, New York, NY 10032 ek2191@columbia.edu)

The *in vivo* cavitation response associated with blood-brain barrier (BBB) opening as induced by transcranial focused ultrasound (FUS) and microbubbles was studied in order to better identify the underlying mechanism. A cylindrically focused hydrophone, confocal with the FUS transducer, was used as a passive cavitation detector (PCD) to identify the threshold of inertial cavitation (IC) in the presence of definity microbubbles. After definity was injected intravenously, pulsed FUS, with parameters previously shown to generate opening, was applied (frequency: 1.525 MHz, peak-rarefactional pressure: 0.30–0.60 MPa, duty cycle: 20%, PRF: 10 Hz, duration: 1 min) on the right hippocampus of five mice *in vivo* through their intact skin and skull. T1-weighted MRI was used to verify BBB opening. A spectrogram was generated at each pressure in order to detect the IC onset and duration. The IC threshold detected was 0.60 MPa and mainly occurred during the first 50 cycles. Harmonics were also detected during BBB opening starting at 0.30 MPa indicating stable cavitation occurrence. In conclusion, stable and inertial cavitation could be detected *in vivo* without craniotomy and IC is not required for BBB opening. [This work was supported by NIH R21EY018505, NIH R01EB009014 and NSF CAREER 0644713.

3:00—3:30 Break

3:30

1pBB9. Cooperative radiation and losses in bubble clusters. Derek C. Thomas, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

In existing models for single bubble dynamics it is necessary to account for radiation damping, a consequence of fluid compressibility. A similar correction is necessary when modeling systems of coupled bubbles. The coupling alters the collective dynamics and therefore the acoustic power radiated. In the linear approximation and for compact clusters in which the bubbles pulsate in phase, the radiation damping per bubble increases in proportion to the number N of bubbles in the cluster, and the acoustic power is N times greater than is radiated by the cluster in the absence of bubble interaction. The latter effect is relevant to passive detection of cavitation noise as an indicator of the onset and degree of cavitation in therapeutic applications such as shock wave lithotripsy and high-intensity focused ultrasound. The effect that collective radiation damping has on the dynamics of several simple systems is considered for both small and large pulsations, and the relative magnitude of the effect is compared with damping due to viscosity and thermal conductivity. [Work supported by the ARL:UT McKinney Fellowship in Acoustics and NIH DK070618.]

3:45

1pBB10. Effects of quadrupole and octupole modes on coupled nonlinear bubble interactions. Eru Kurihara, Todd A. Hay, Jason P. Kurtz, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

Photographs of bubble cluster dynamics have shown that under sufficiently intense acoustic excitation, neighboring bubbles may undergo shape deformation and even form opposing liquid jets. Jetting is one of the chief mechanisms thought responsible for cavitation-induced erosion. Therefore, a better understanding of the phenomenon would be beneficial in improving the efficacy of and reducing collateral damage during medical procedures such as shock wave lithotripsy. Previous efforts to model aspherical collapse of a bubble near an interface have utilized purely numerical methods. In this presentation we derive dynamical equations for the coupled motion of two translating, aspherical gas bubbles. Deformation of the initially spherical

bubble surfaces is taken into account by including quadrupole and octupole spherical harmonics in the boundary conditions on each bubble. For an axisymmetric geometry, this results in a set of eight coupled second-order ordinary differential equations that may be integrated numerically in time. The effect of surface tension and shear viscosity on quadrupole and octupole contributions is included in a consistent manner. Simulation results agree qualitatively with measurements and predict the formation of opposing liquid jets for bubbles which are sufficiently close to one another. [Work supported by NIH DK070618.]

4:00

1pBB11. A new acoustic lens design for electromagnetic shock wave lithotripters. Nathan Smith, W. Neal Simmons, Georgy Sankin, and Pei Zhong (Dept. of Mech. Eng. and Mat. Sci., Duke Univ., Durham, NC 27708)

The third-generation electromagnetic (EM) shock wave lithotripters often have narrow focal width and high-peak pressure compared to the original HM-3. In addition, the pressure waveform produced by a typical EM lithotripter has a secondary compressive wave that suppresses lithotripter pulse induced cavitation, which is an important mechanism for stone comminution. These features have been contributed to the reduced effectiveness of the modern EM lithotripters. To overcome these drawbacks, we have designed a new acoustic lens for the Siemens Modularis EM lithotripter that produces an idealized pressure waveform similar to the HM-3 with broad focal width and low-peak pressure. At acoustic pulse energy of 53 mJ, the new lens design enlarges the -6 dB focal width of the Modularis by 47% while significantly reducing the second compressive wave throughout its focal plane. After 2000 shocks, *in vitro* comminution produced by the original and new lens designs are 100% and 99% at the lithotripter focus, and (49 .7$\pm$15.0)% and (75.3$\pm$8.74)% ($p < 0.001$) at 10 mm off axis, respectively. Corresponding values *in vivo* are (63.4$\pm$17.4)% and (83.3$\pm$8.5)% ($p < 0.01$), demonstrating the significant performance improvement provided by the new lens design. [Work supported by NIH Grant Nos. R01DK052985 and S10RR016802.]

4:15

1pBB12. Gas content of the medium surrounding a stone has a significant effect on the efficiency of stone breakage in shock wave lithotripsy. Yuri A. Pishchalnikov (Dynaflow, Inc., 10621-J Iron Bridge Rd., Jessup, MD 20794, yurapish@gmail.com), Anthony J. Zancanaro, James C. Williams, Jr., and James A. McAteer (Indiana Univ., Indianapolis, IN, 46202-5120)

In shock wave lithotripsy inertial collapse of cavitation bubbles produces daughter microbubbles [Y. A. Pishchalnikov *et al.*, BJU Int. **102**, 1681–1686 (2008)]. This will cause proliferation of bubbles only if daughter microbubbles survive until the arrival of the next shock wave (SW). While the time interval between SWs is determined by the firing rate of the lithotripter, the dissolution time of microbubbles depends on concentration of dissolved gas in the water. Thus, the same firing rate may produce different cavitation fields depending on the gas concentration. We assessed how the concentration of dissolved gas in the water tank affected SWs and the efficiency of stone comminution. More numerous bubbles generated in nondegassed water caused a dramatic reduction in the trailing negative-pressure component of the SWs, while the leading positive-pressure phase remained almost unchanged. It required 3.5 times more SWs to break stones in nondegassed water than in water degassed to 30% ($P<0.001$). This suggests that changes in temporal profiles of the SWs associated with cavitation are responsible for reduced efficiency of stone comminution observed in dense cavitation fields, regardless of whether these denser cavitation clouds were generated by using a faster rate or higher gas content. [Work supported by NIH-DK43881.]

4:30

1pBB13. Hydrophone array for instantaneous measurement of lithotripter fields. Jonathan M. Kracht (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), Jeffrey A. Ketterling (Riverside Res. Inst., New York, NY 10038), and Robin O. Cleveland (Boston Univ., Boston, MA 02215)

Electrohydraulic lithotripters employ a spark placed within a hemiellipsoidal reflector to generate shock waves (SWs). The shot-to-shot jitter in the spark location results in variability in acoustic measurements, e.g., 50% variation in peak positive pressure. Field measurements with a single-element hydrophone therefore only provide average field properties over many SWs. The ability to obtain an instantaneous "snapshot" of the sound field would have broad implications for advancing the understanding of how lithotripters fragment stones and damage kidney tissue. Here linear hydrophone arrays consisting of 20 elements were created by bonding a piezopolymer film to copper-clad polyimide. An array pattern was etched on the copper to provide individual connections to the 20 elements. Failure testing of the hydrophone array was conducted in a piezoelectric lithotripter that generated 60 MPa peak pressure. The sensitivity of the hydrophone remained relatively constant through the first 500 SWs and then gradually decreasing over a 500–1000 SW period. This reduction was attributed to damage from cavitation activity and was confirmed by visual inspection after testing. A new hydrophone was developed and results from measurements of an electrohydraulic lithotripter will be reported. [Work supported by the National Institutes of Health (DK081119 and DK43881).]

4:45

1pBB14. Pocket-sized ultrasonic surgical and rehabilitation solutions: From the lab bench to clinical trials. George K. Lewis, Jr., William L. Olbricht, and Peter Henderson (Dept. of Biomedical Eng., Cornell Univ., 108 Olin Hall, Ithaca, NY 14853, george@cornellbme.com)

The use of ultrasound in medicine has quickly evolved in the last decade as a therapeutic modality and means to treat disease independently or in combination with other therapeutic regimes such as chemotherapy. The interaction of ultrasonic waves with soft mammalian tissues has the innate impact of enhancing tissue permeability to therapeutics, increasing local blood flow at low-acoustic powers, and causing tissue necrosis and cell death at higher-energy levels. We have developed multiple pocket-sized (4$\times$2$\times$1 in.) lithium battery powered therapeutic ultrasound units based off of our novel ultra-efficient electrical amplifier design that produces acoustic energies for low- and high-power applications (1–130 W). The portable ultrasound technology has been used in our laboratory to enhance ten-fold the delivery and efficacy of convection enhanced drug delivery to treat brain gliomas *in vitro* and *in vivo*. Our clinical collaborates at Weill Cornell Medical Hospital use our devices at high powers to non-invasively cauterize veins as a novel method of varicose vein treatment. Our clinical collaborators provide new avenues of ultrasound research and critical user feedback on the technology to aid refinements in its implementation in hand-held acoustic applications.

Session 1pMU

Musical Acoustics: Homemade Musical Instruments for Teaching Acoustics

Thomas D. Rossing, Cochair
Stanford Univ., CCRMA Dept. of Music, Stanford, CA 94305

Preston S. Wilson, Cochair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin, TX 78712-0292

Andrew C. H. Morrison, Cochair
DePaul Univ., Physics Dept., 2219 Kenmore Dr., Chicago, IL 60614

Chair's Introduction—1:00

Invited Papers

1:05

1pMU1. Learning by playing: Musical acoustics of found objects. Laurie McNeil (Dept. of Phys. and Astronomy, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3255, mcneil@physics.unc.edu)

"Teaching by telling" has been shown to be ineffective in many physics contexts, and acoustics is no exception. In my 10 years of teaching, a freshman seminar "The Interplay of Music and Physics," about the physics of musical instruments, I have learned that there are some concepts that students learn only by making their own instruments and discovering how the success of the instruments depends on them. My colleague Professor Brent Wissick of our Department of Music and I have team-taught this class to college students of a broad range of backgrounds and majors. The students design and build stringed and wind instruments out of whatever they can find lying around their dorm rooms. At the end of the semester, they give a concert of music of their own composition, played on the instruments they have created. In this presentation, I will discuss what we have learned about how such activities teach the students acoustical principles in ways that more conventional means do not accomplish.

1:30

1pMU2. Homemade "woodwind" and "brass" instruments. Stephen Luzader (59 Centennial St., Frostburg, MD 21532, sluzader @frostburg.edu)

PVC pipe is an ideal material for the construction of wind instruments, both "woodwind" (using a reed) or "brass" (using the lips). Instruments can be made using a combination of PVC construction along with parts of actual musical instruments, such as a clarinet mouthpiece and reed or a trumpet mouthpiece, or they can be manufactured entirely from PVC and other materials. For panpipes, the lengths of the tubes are easily determined by choosing a standard pitch frequency and the basic interval between tones (e.g., 440 Hz and equal temperament) and then using basic physics to calculate the required tube lengths. For instruments using side holes, the locations of the holes can be determined by choosing a standard pitch frequency and interval and then calculating the required distances of the holes from the driven end of the pipe. The basic theory will be reviewed, and several different instruments will be demonstrated.

1:55

1pMU3. Homemade musical instruments for the classroom. Thomas J. Senior (355 Dell Ln., Highland Park, IL 60035-5310)

Pan pipes produced using drinking straws provide a very inexpensive way for students to study the nature of resonant columns and musical scales. A slide whistle is also easily made using the straws. Most students easily learn to produce tones from the pan pipes/slide whistle. For situations where too many loud sounds would be undesirable, the "Palm Pipe" offers a similar learning experience. The palm pipe is a single length of PVC pipe that can be played by striking the palm on the end of the tube to produce a tuned percussive sound. Methods for the production of slide whistles, pan pipes, and palm pipes will be described and demonstrated in this presentation. A few other "musical" devices will be shown as time permits, including a stadium horn and a thunder drum.

2:20

1pMU4. Homemade musical instruments. Ann Sayre Wiseman (284 Huron Ave., Cambridge, MA 02138, ansayre@aol.com)

When the creative arts were dropped from the school curriculum for lack of money in 1972, as Program Director at the Boston Children's Museum Visitor Center (Jamaica Plain), we created low-cost learning by doing workshops for teachers using available materials. Out of this an orchestra of good sounds was born. Good sounds found in the kitchen and junk yard by tapping everything to hear its note. We made excellent conduit-pipe xylophones, drums, horns, strings, and percussion instruments hung from a frame. A low cost musical program any teacher could include in the curriculum, involving math, history, and research, the kind of learning you never forget when making useful things by hand with your own initiative. These recipes became Making Things Handbook of Creative Dis-

coveries, winning awards, and freely photocopied by teachers who wrote this book is my classroom bible. The handmade orchestra became another book: Making Musical Things and in 2002 Storybook Press republished it as Making Music, including John Langstaff's Inovative Clock Orchestra and the Art of Conducting. As an Expressive Therapist in 2005 with International exchange groups, we took these workshops abroad to show students and medical professionals the healing powers of the creative process.

2:35

1pMU5. The homemade didgeridoo: An ancient acoustic resonator. Chad A. Greene (Dept. Mech. Eng. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

The didgeridoo is an ancient Aboriginal instrument traditionally crafted from Eucalyptus trees which have been hollowed out by termites. The instrument is an excellent example of a simple one-dimensional acoustic resonator with a fundamental resonance corresponding to one-half wavelength. Today, the ubiquity and low cost of PVC pipes make didgeridoos inexpensive, easy to find, and trivial to construct from materials found at any hardware store. Students will have fun learning to make sounds with the instrument—perhaps they will even learn the technique of circular breathing—all while exploring fundamental principles of acoustics. As further investigation students can hum as they play, resulting in "beats, the speed of which can be adjusted by changing the pitch of the hum.

2:50

1pMU6. Homemade digital musical instruments. Edgar Berdahl, Wendy Ju, and Julius O. Smith, III (Dept. of Music, CCRMA, Stanford Univ., Stanford, CA 94305, eberdahl@ccrma.stanford.edu)

Homemade musical instruments are wonderful tools for getting students interested in learning about acoustics. Homemade *digital* musical instruments can similarly engage students, provoking them to learn more about the modern world surrounding them. In addition, the tools for developing homemade digital musical instruments are now becoming so easy to access that students can benefit from them too. The Arduino tool chain is introduced for interfacing sensing circuits with the computer over a USB port. Modern sensors are so simple that only a basic understanding of the voltage divider is required for building many useful sensing circuits. Synthesizing sound according to incoming data from the sensors is equally straightforward due to graphical environments commonly employed for sound synthesis programming. While many sound synthesis models are available, physical models of acoustic musical instruments are especially intriguing because students can relate to interacting with physical objects. Step-by-step instructions are provided for ordering parts and creating a simple but easily extensible homemade digital musical instrument that triggers sounds from a physical model of a vibrating string.

3:05—3:15 Break

3:15

1pMU7. Pedagogical possibilities: Some ideas for instrument making in the classroom. Bart Hopkin (P.O. Box 421, Point Reyes Station, CA 94956, emi@windworld.com)

As a brief introduction, historical etchings of a few intriguing 19th century contraptions for classroom demonstrations of acoustic principles will be shown. Following that, several classroom-buildable musical instruments will be described. The emphasis will be on instruments that effectively illustrate acoustic or mathematical principles. They will also be fun to make and play, and practical in a classroom context, requiring a minimum of special tools, skills, or materials. Topics to be illustrated through hands-on instrument making may include relationships between frequency, wavelength and pitch, harmonic and inharmonic overtones, relationships between drivers, radiating surfaces and atmosphere, linear and inverse-squared functions, and more.

Contributed Papers

3:30

1pMU8. Building a musical instrument as a laboratory project in a short-term course. Andrew C. Morrison (Dept. of Phys., DePaul Univ., 2219 N. Kenmore, Chicago, IL 60614-3504)

In a month-long short course titled "Sound, Music, and Hearing" students were required to design and build their own instruments as a final laboratory project. The assignment requires that the instrument be able to play a scale of a minimum of five notes recognizable to the ear of the instructor. Students were allowed to work in pairs, or alone on the instrument laboratory. A review of strategies for overcoming the challenges of including an instrument building laboratory in a short-term course designed for general education credit is presented.

3:45

1pMU9. America's next top physics rockers: A project based learning unit in which students build, perform, and explain homemade instruments. Charles H. Sabatier (Mount Vernon High School, 8515 Old Mount Vernon Rd., Alexandria, VA 22309, charles.sabatier@fcps.edu)

Project based learning requires that students engage with content in order to be successful in completing a project. This teaching method creates a need-to-know which motivates students to learn more so than the traditional physics unit. Knowledge and skills outside of the physics content such as preparing presentations, using tools and technology, and utilizing different writing styles are required to be successful in the project. Making the project authentic is necessary to truly motivate students to exceed expectations. In this particular project based learning unit, students have built their own musical instruments, performed a musical piece in a small group, and taught others about how their instruments work individually and as an ensemble. In completing this project, students learn about vibrations, resonance, amplitude, frequency, wavelength, and how the characteristics of their instruments affect these physics principles. In order to make this project truly authentic, the top students will be selected to perform and explain their instruments as part of this talk. All students participating in this project (pending funding) will attend the session and support their classmates in this presentation. If funding is not available, video clips of student performances and explanations will be substituted.

Invited Paper

4:00

1pMU10. Homemade musical instruments from Westtown school. Barry, H. Feierman (Dept. of Phys., Westtown School, West Chester, PA 19382)

Westtown School has implemented a team-taught collaborative class in high school featuring the Music and Science Departments. A new course, The Physics of Sound/Music, aims to teach students about the physics of sound waves and music theory. The course is open to 11th/12th graders and is team-taught by a science teacher and a music teacher. Objective: Each student will demonstrate his/her knowledge through the design and construction of a new, unique musical instrument,write a musical composition to feature the instrument, and play that piece in a student recital. This course is an elective open to students who have already completed chemistry and physics, and appeals to students who would like a creative challenge in terms of both design and construction of a musical instrument. This paper will be to describe the nature of the course, the experience of its two teachers, and to allow some of the students in the class to demonstrate their home-made instruments and talk about their experience. [Westtown School is a K-12 Quaker co-ed boarding school near Philadelphia.]

4:15—5:00 Demos

MONDAY AFTERNOON, 19 APRIL 2010 GRAND BALLROOM I/II, 1:00 TO 3:15 P.M.

Session 1pNCa

NOISE-CON and Noise: Community Noise

Courtney B. Burroughs, Chair
Noise Control Engineering, Inc., 1241 Smithfield St., State College, PA 16801

Contributed Papers

1:00

1pNCa1. Modeling with the use of the commercial software package CADNAA (Computer-Aided Noise Abatement) to estimate the probability of awakening associated with train horns. Stephen Szulecki (The Noise Consultancy, LLC, 309 Van Neste Rd. Flemington, NJ 08822, steve-noiseconsultancy@comcast.net), Eric Zwerling, Craig Anderson, and Barbara Turpin (Rutgers Univ. Noise Tech. Assistance Ctr.)

The Rutgers University Noise Technical Assistance Center conducted a long-term acoustical study to assess the noise exposure of residents in Teaneck, NJ to noise associated with idling freight trains. One of the noise sources evaluated was idle-related horn usage, at locations other than grade crossings. Trains on an adjacent track overtaking an idling locomotive regularly blow their horn as they approach and pass an idling train. The primary impact on the community due to this type of horn use is sleep disturbance. In order to evaluate this impact, an analysis was performed using CADNAA acoustical modeling software, in an innovative manner, to evaluate the probability of awakening (POA) as a result of idle-related horn usage based on ANSI/ASAS-12.9-2008/Part 6, a standard for estimating awakenings. The standard presents a series of formulas that calculate the POA as a function of a sound exposure level of a single or series of noise events. Other factors influencing the POA are time since retiring and whether the sound is new or existing. The model output is a graphic of equal-probability contours showing the POA at least once throughout the night.

1:15

1pNCa2. Transit noise and community choices. Ahmed El-Aassar (5801 Grayson Rd., Harrisburg, PA 17111, aelaassar@envact.com) and David Dubbink (Noise Management Inst.)

Transit innovations and expansions are a central part of initiatives to construct more livable communities. As part of the National Environmental Policy Act process, public involvement is required and public feedback and support are essential for projects to move forward. Generally, accessibility improvements are welcomed by the public, but the potential for increased noise is often met with opposition—and concerns about transit noise can stifle initiatives. One way to build community acceptance is to accurately compare transit options. The authors explain how community groups and decision makers in Florida responded to a presentation system providing direct acoustic experience of over a dozen transit modes, from light rail guideway transit to commuter railways. The goal of the acoustic presentation was to compare existing freight train sound levels to those of an array of transit technologies. The audio-visual presentation included interactive videos and computer animations, allowing proposed transit modes to be heard at different distances and in outdoor or indoor settings.

1:30

1pNCa3. Contemporary consulting practice in California's central coast. David Lord (45dB.com, 299 Albert Dr., San Luis Obispo, CA 93405, dl@45db.com)

The resolution of technical, theoretical, and practical issues for a small consultancy in a remote location brings better results and greater satisfaction to all involved. This paper will describe the range of consulting projects encountered in a typical year and will briefly cover the technical tools used and their application to modest-scale residential and commercial development in the Central Coast of California. The acoustics consultant's responsibility and influence on project development require technical expertise, negotiation, and tailored explanations-on-demand of principles and details to clients, design professionals, builders, and planners as well as interested and sometimes skeptical community groups.

1:45

1pNCa4. A case study in cooperation: A gravel pit and its community. Edward Duncan and Kenneth Kaliski (Resource Systems Group, 55 Railroad Row, White River Junction, VT 05001, eduncan@rsginc.com)

It is often the case in the permitting and operation of gravel extraction and crushed stone quarries that aggregate companies and the local community are pitted against one another. This paper presents a case study of an existing gravel pit that came together with the community to work through concerns about the existing operations and a proposed expansion. The topics

of discussion include the interesting dynamics involved in the consulting process, resulting mitigation to meet local standards and community concerns, conflicts of interests, and the local permitting process. The cooperative process used is presented in contrast to traditional systems using case studies of projects that have undergone permitting using more adversarial process models.

2:00

1pNCa5. Evaluation of crowd noise in Beaver Stadium during Penn State football games. Andrew Barnard, Stephen Hambric (ARL/Penn State, P.O. Box 30, State College, PA 16804, arb279@psu.edu), and Steve Porter (Head Acoust.)

Sound pressure level measurements were made at Penn State (PSU) football games at Beaver Stadium, on the PSU campus, in Fall 2009. Twelve sound level meters were positioned around the field, on the press box, on the luxury suites, and outside the stadium to capture the characteristics of the crowd noise. The loudest portion of the sideline area was determined along with the maximum equivalent *A*-weighted sound pressure level. Additionally a HEAD VISOR acoustic array was utilized in conjunction with a portable white noise generator in order to determine which portions of the student section were generating the majority of the sound pressure on the field level. The results indicate which sections of students are able to most effectively project their voices onto the field.

2:15

1pNCa6. A three-stage approach to understanding listener perception of weapon signature for small arms. Jeremy Gaston and Tomasz Letowski (Dept. of the Army US Army Res. Lab., US Army Res. Lab., Aberdeen Proving Ground, MD 21005-5066)

The ability to identify and interpret impulse sounds from small arms weapon fire is a very important element of soldier's situational awareness that is critically needed to avoid potential danger. For example, differentiation of hostile from friendly weapon fire can indicate the need for increased vigilance and can reveal the approximate location of an enemy element. In addition, estimation of weapon size can be used to infer the resources of an enemy element. Despite the potential operational importance of these sounds, little is known about listener perception of impulse weapon sounds. The present work investigates listener ability to differentiate signatures of various small arms based on high-quality recordings of single-shot impulse sounds. Following the three-stage approach developed by Pastore *et al.* (2008) for studying natural sound perception, the interrelationships between weapon source properties, the physical attributes of the sounds produced and listener perception are evaluated for a number of small arms weapons. The mappings of these relationships provide the basis for developing training techniques for improving listener weapon signature identification ability.

2:30

1pNCa7. Design and test of a manufacturing building to meet a 40 dB from above 300 Hz up property line night noise limit. Angelo Campanella (Campanella Assoc., 3201 Ridgewood Dr., Hilliard, OH 43026, a.campanella@att.net)

A new one-story manufacturing plant was designed to operate around the clock 7 days a week. Its south property line adjoined existing residential lots. The existing noise ordinance required that the property line noise from the plant shall not exceed 40 dB at all frequencies above 300 Hz at night. The city had already erected an eight barrier mound near that line. Positions of plant air handling HVAC systems were revised to minimize sound emission to the south. Barriers were designed for the outdoor HVAC condensers. The building wall near noisy plant equipment was provided extra attenuation. On commissioning, it was found that a fan needed on the south facade was running above design speed and was readjusted. Sound at the property line was then so quiet compared to the environment that additional monitoring was conducted halfway between the plant wall and the property line. Measurement at the property line was made in 1/3 octave bands, and then parsed to approximate the old octave band limits format to demonstrate compliance.

2:45

1pNCa8. Limiting technology for modeling community noise from petrochemical facilities. Frank Brittain (Bechtel, Corp., 2255 Peavine Valley Rd., Reno, NV 89523, fhbritta@bechtel.com)

Modeling community noise is a vital part of designing any petrochemical, power, or industrial facility to meet community noise limits. Increasingly ray tracing, which is state-of-the-art software, is being used. Limitations of ray tracing can be categorized as those arising from the commercially available software (including ISO 9613 usually used for calculating outdoor propagation) and those related to input to the software. The former was covered by Brittain and Hale, Noise-Con 2008. The latter is covered in this paper. There are many technologies, which are needed to model community noise that are weak or for which data are not readily available. If the modeler understands where the technology is weak, then more effective models can be developed, and inaccuracies in the result can be better interpreted. Weak technologies or lack of data includes the following: modeling air coolers and cooling towers, modeling ducts and pipe racks, IL of pipeline and duct lagging, TL of building walls, noise radiated by building walls, source directivity, and—last but not least—getting accurate noise source data. Most of these become more critical at low frequencies. Each of these and their impact on design will be discussed.

3:00

1pNCa9. Creation of combined noise maps for noise prediction in petrochemical plants. Arindam Ghosh (KT-2425C, 601 Jefferson, Houston, TX 77002, arindam.ghosh@kbr.com)

Design of petrochemical plants involves producing detailed noise models with predicted noise contours at fine grids such as 1 m. However, background noise data collected during pre-construction environmental impact assessments (EIAs) are typically gathered at a few selected points. Since it is impractical for all the measurements to match the prediction grid, the measured data must be interpolated. This paper discusses how the survey grid is selected and interpolation optimized to meet the overall project noise control design basis and the post-start-up noise survey needs. The strategy varies based on the existing operational situation (brown-field or green-field), the type of land-use (rural or urban), the nature of terrain (flat or hilly), and the planned plant use (process area, loading and storage area, or administrative buildings). Finally, the unique needs for petrochemical process plants (as opposed to general city planning) for long term and area-wide background noise measurements are outlined.

Session 1pNCb

NOISE-CON and Noise: Industrial and Power Plant Noise

Frank H. Brittain, Chair
2255 Peavine Valley Rd., Reno, NV 89523

Contributed Papers

3:30

1pNCb1. Petrol-chemical plant noise models: Noise surveys, model development and validation. Hans Forschner, Sr. and Jim Steedman (Navcon Eng. Network, 701 W. Las Palmas Dr., Fullerton, CA 92835 Forschner@navcon.com)

Developments in computer modeling & presentation software has changed how acoustical engineers approach noise planning, mitigation, trending projects. The ability to quickly and accurately develop a three dimensional (3-D) noise model of a project area and/or to create noise contours from measured data allow for a more rigorous analysis in a shorter period of time while maintaining a level of consistency with current acoustical engineering standards and practices. This paper describes the procedures and methods for generating 3-D computer noise models, validating noise models, and graphically comparing measured and predicted noise levels.

3:45

1pNCb2. Noise prediction for industrial facilities. Wolfgang Probst (DataKustik GmbH, Gewerbering 5, 86926 Greifenberg, Germany, info @datakustik.com)

The methodologies used to predict the noise caused by industrial facilities are presented and ranked. There are a lot more scientifically based methods that use certain approximations to solve or apply the wave equation in a layered atmosphere. These methods are usable to predict even meteorological influences with long distance propagation but can hardly be used with complex source and surface geometries for receivers nearby. This latter case is the preferred application of engineering models using simple geometric ray approximations. The techniques to use it with complex and extended sources are presented. This includes the use of parametrized source description for technical sources such as motors, gears, pipes, and cooling towers and numerical aspects such as projection or angle scanning. It is shown how an intelligent software technique like the "object tree" supports the generation of variants of a plant including different steps of noise reduction measures.

4:00

1pNCb3. Ray-tracing and spreadsheet modeling of power plant noise. Marlund E. Hale (Adv. Eng. Acoust., Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, noisedoc@aol.com)

Community noise limits are often imposed on power plant designs. Many power plant sites are planned for rural locations where the residences are relatively close by or at some distance away. During the preliminary design of power plants, design options, along with any predicted environmental impacts and needed mitigation measures, are often discussed with appropriate regulatory agencies and local residents. A design phase noise model can be used to demonstrate the compliance status with local noise regulations. Several proprietary noise modeling methods have been developed by engineering and design-build companies. Spreadsheet noise models are often used to obtain a preliminary look at potential plant noise impacts. More sophisticated, commercially available, ray-tracing software is also available, but at a large initial cost in both time and money. Such modeling software allows one to model a large array of conditions and factors affecting noise propagation that are not possible using simple spreadsheet modeling. This paper compares the modeling results by both methods for two very different power plant sites, one surrounded by flat open terrain and the other located in a small valley surrounded by hilly forested terrain. Measured property line and community noise data are presented along with the modeling results.

4:15

1pNCb4. Anatomy of a quiet power plant: A case study. Jeff Szymanski (Black & Veatch, 11401 Lamar Ave., Overland Park, KS 66211, szymanskijd@bv.com)

Details of noise control design are presented for a recent power plant expansion project that involved the addition of heat recovery steam generation to a gas-fired power plant in the US. The project, including its expansion, was subject to a 45 dBA nighttime property boundary sound level limit. The development and coordination of noise control design for large and small equipment packages, some of which would be located within several meters of the property boundary, are discussed in detail. Sound level measurements are presented and discussed, particularly with regard to the challenges involved with conducting sound level measurements in the presence of ambient sound levels that often approached the allowable noise limit.

4:30

1pNCb5. Pipe noise prediction and sound insulation design by using modelling and field measurements. Giuseppe Squadrone (Tecnimont S.p .A., viale Monte Grappa, 3 20124 Milano Italy, g.squadrone@tecnimont.it)

Since engineering companies usually develop industrial complex by combining and connecting among them many static and rotating equipment with piping, the noise calculation and model of the latter are necessary when carrying out acoustic design to comply with occupational and environmental noise limits. According to the set of available input data, different design procedures can be followed both to model pipe noise and to design sound reduction systems, if needed. Examples of methods for piping noise modeling and insulation design, starting from inside or outside pipe noise levels given by equipment supplier or from in field measured noise or vibration levels, are here shown, analyzed, and discussed. Specifically, the paper describes case studies of steel piping system relevant to installation usually present in a generic industrial plant: a steam pressure reduction station, constituted by control valves and desuperheating devices; a large cooling water pump; and a feeding cooling water station where, accidentally, a regulation valve was excessively noisy. In-plant noise measurements after start up, for methodology reliability evaluation, are reported and discussed too.

Session 1pNCc

NOISE-CON and Noise: Mufflers and Silencers

Mark Storm, Chair
URS Corp., 1615 Murray Canyon Rd., Ste. 1000, San Diego, CA 92108

Contributed Papers

1:00

1pNCc1. Suggested modifications to the American Society for Testing and Materials E-477. Jerry Lilly (JGL Acoust., Inc., 5266 NW Village Park Dr., Issaquah, WA 98027, jerry@jglacoustics.com)

ASTM E-477 is the current standard test method used in North America for evaluating the acoustic and aerodynamic performance of duct silencers. The test method uses, at a minimum, a source chamber, a test duct (that includes a test specimen), and a reverberation chamber. If testing under flow is desired, the system also requires a fan chamber that can provided quiet airflow through the test duct. The results of the test method yield silencer insertion loss, self-generated noise, both as a function of acoustic frequency and flow velocity, and pressure drop. There are several problems with the current test method, some that are being addressed by the ASTM working group that is responsible for the development and maintenance of this standard. This paper will highlight the most significant problems associated with the current standard and propose modifications to the standard that should, if adopted, improve the intra-laboratory reproducibility of the test results.

1:15

1pNCc2. Model for the design of a hydraulic silencer with dispersive liner. Ken Marek, Nick Earnhart, and Ken Cunefare (George W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr., J. Erskine Love Bldg., Grad Box 334, Atlanta, GA 30332, ken.marek@gatech.edu)

An in-line hydraulic silencer with a dispersive lining material shows promise for effective fluid noise suppression. For this design, a rigid outer shell encases an annular dispersive medium, where fluid flow is straight through the annulus. A model is developed to predict input port impedance and acoustic transmission loss of the hydraulic silencer. These predictions are used to develop a set of design parameters for the liner, including material properties and geometric dimensions. Design parameters can then be optimized for a given set of working conditions and frequency ranges of interest. Model development is discussed along with comparisons to experimental data.

1:30

1pNCc3. Evaluation of hydraulic silencers. Nick Earnhart, Ken Marek, and Ken Cunefare (Georgia Inst. of Technol., 771 Ferst Dr., Grad Box 261, Atlanta, GA 30332, nick.earnhart@gatech.edu)

Hydraulic silencers are devices used to mitigate fluid-borne noise in fluid power systems. Fluid-borne noise may be produced by positive-displacement pumps and interacts with system components to generate vibration and air-borne noise. It is of interest to quantify the performance of silencers with regard to the input reflection coefficient and transmission loss. A test rig has been constructed that uses a six-sensor method to determine these quantities. Construction and methodology will be presented along with results for apparent transmission loss comparing a commercially available silencer to a novel prototype.

1:45

1pNCc4. Optimal silencer baffle positioning within an air handling unit. Mark Storm (URS Corp., 1615 Murray Canyon Rd., Ste. 1000, San Diego, CA 92108, mark_storm@urscorp.com)

A stationary bank of rectangular modular sound traps or free-standing sound attenuating parallel baffle "columns" installed within a central station air handling unit (AHU) is a common technique for controlling fan noise in a ducted heating, ventilating, or air-conditioning system. Recent design developments by a leading custom AHU manufacturer offer the ability to move these columns within the AHU, so that they can be positioned close to a noise source (e.g., a plenum fan) during operation and pushed away in order to allow AHU component inspection and maintenance. While the columns have a fixed shape and set of features (perforated facing, sound absorbing fill, etc.), this study examines their sound attenuating performance relative to position variance from the source (i.e., fan noise) and the AHU inlet or outlet port at which desired sound levels are often specified. The objective, should such column position variance has meaningful influence on net transmission loss or noise reduction for the AHU, is to estimate optimal geometry(ies) than can be tested and subsequently established as guidance for superior AHU noise control design.

2:00

1pNCc5. Mine ventillation silencer. Jonathan Weinstein (Industrial Acoust. Co., 1160 Commerce Ave., Bronx, NY 10462, jweinstein @industrialacoustics.com)

Mine exhaust ventilation fans provide essential fresh air. The extracted air from the mine is often moist and laden with dust, so that conventional silencers would very quickly clog as the wet dust forms on silencer baffles and blocks the perforations in the baffle skins. This paper presents a case history of the solution to this problem where tuned cavity silencer baffles were installed in an existing exhaust evase of a mine ventilation fan. Sound attenuation is achieved using tuned cavities which are resistant to blockage caused by the wet dust. This performance was successfully demonstrated with a half scale model tested in a laboratory wind tunnel.

2:15

1pNCc6. A true to life application for flow driven resonators: Industrial chiller cooling fan noise reduction. Lee J. Gorny and Gary H. Koopmann (The Pennsylvania State Universtiy, 157 Hammond Bldg., Univ. Park, PA 16802, ljg138@psu.edu)

A project has been conducted to verify the efficacy of tunable resonators in reducing the tonal noise output from an industrial chiller's cooling fan. The primary goals of this work were to demonstrate resonator source robustness for attenuating blade tone noise generated by fixed speed fans, radiating primarily in a single direction, that have non-ideal, un-ducted flow conditions. BPF tones radiated by the fan are reduced by as much as 14 dB in the flow stream and 10 dB outside, demonstrating the effectiveness of resonators for realistic applications. A manually tunable fixed mouth location resonator configuration was designed and developed as a first step in testing resonator effectiveness for quieting production chillers operating at a fixed speed and steady loading. The test enclosure was developed to simulate the fan's environment for a production device, mimicking the convoluted inlet flow, exit through a radiator. Changes in loading were accomplished through adjustment of a flow control door. Additional flow obstructions inside of the enclosure added and removed to demonstrate their effect on overall tonal noise radiation and to test resonator effectiveness in reducing various interaction noise sources.

2:30

1pNCc7. Dissipative diesel engine exhaust muffler featuring sintered metal fiber as acoustical fill liner. Mark Storm (X Sound Technologies LLC, 4251 E. Thomas Rd., Phoenix, AZ 85018, mstorm@xsoundtech.com)

Conventional exhaust muffler design for diesel truck applications often involves a combination of dissipative and reactive elements in order to provide generally broadband sound attenuation for the purpose of suppressing engine exhaust noise during normal operation or engine braking (i.e., venting compressed air through the exhaust system). The reactive portions of the muffler are usually present to provide low-frequency spectrum transmission loss that cannot readily be achieved with classical perforated-n'-packing dissipative design. However, unlike standard perforated metal liners that are used to separate the exhaust gas stream from the muffler's surrounding volume of acoustically absorbent fill, liner materials such as micro-perforate and sintered metal fiber (SMF) metals offer high porosity and high-flow resistance that should improve low-frequency dissipative sound absorption. This study contrasts the modeled insertion loss, transmission loss, and noise reduction performance of a dissipative-only muffler design featuring an SMF liner with that of a popular conventional dissipative-reactive combination muffler design. Discovered performance advantages and disadvantages should reveal pathways for subsequent muffler design development that utilizes advanced liner materials such as micro-perforates and SMF.

2:45

1pNCc8. Noise reduction under the hood of a portable air-purifying respirator system by means of a muffler insert. Robert Marsh (Polymer Technologies, 420 Corporate Blvd., Newark, DE 19702, bobm@polytechinc.com)

Speech intelligibility is probably never more important than in emergency situations. Recently a manufacturer of a portable air-purifying respirator system requested assistance in lowering the noise level that users were encountering under the hood. In operation, the noise level and sound quality in the hood were reported to be objectionable and concern about the ability of personnel to communicate effectively was expressed. Typically, choosing quieter components would be the first step in evaluating various acoustical treatment packages. However, since the components of the blower exposed to the air path were already NIOSH approved, choosing an alternate blower was not a viable option. Polymer Technologies embarked on a study that evaluated the blower and the paths to the receiver, successfully providing attenuation options that could be easily implemented. Prototype testing proved that substantial acoustical attenuation could be achieved by utilizing an in-line muffler lined with acoustical absorption material. The maximum benefit achieved was in excess of a 14 dBA reduction in the overall sound pressure level under the hood at ear level and resulted in what was considered to be a more "acceptable" sound.

MONDAY AFTERNOON, 19 APRIL 2010 GRAND BALLROOM III/IV, 3:15 TO 4:30 P.M.

Session 1pNCd

NOISE-CON, Noise, and Structural Acoustics and Vibration: Tire/Pavement Noise

Robert O. Rasmussen, Chair
The Transtec Group Inc., 6111 Balcones Dr., Austin, TX 78731

Contributed Papers

3:15

1pNCd1. Rapid deterioration of sound level benefits for quieter pavements in Washington state based on the on-board sound intensity method. Tim Sexton (Washington State Dept. of Transportation (WSDOT), 15700 Dayton Ave. N, NB82-138 Seattle, WA 98133, sextont@wsdot.wa.gov)

This paper provides detailed descriptions of the three test sections of "quieter" pavement evaluated by the Washington State Department of Transportation (WSDOT) using the on-board sound intensity (OBSI) method and the unique deterioration of those pavements. Within 6 months of installation, OBSI for the test sections were not audibly lower than the control sections. OBSI is the predominant metric used to evaluate the success of quieter pavements in Washington State. Monthly OBSI measurements describe the seasonal variability in measured sound levels and highlight more rapid general sound level increases in the test pavements relative to the control. Asphalt mix compositions and installation procedures used in Washington were similar, or identical, to those used in southern states where the same pavements have been successful. Challenges to the success of all pavements in Washington State include studded tires, frequent freeze-thaw cycles, and cool summers that encourage lower surface temperatures during installation than in southern states. These challenges are particularly significant for open-graded friction course OGFC pavements as evidenced by spikes in OBSI measurements following seasonal weather events. Parallels between OBSI measurements and structural durability measurements are described in detail.

3:30

1pNCd2. The acoustic longevity of various pavements for noise reduction performance based on on-board sound intensity measurements. Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd., South Petaluma, CA 94952, pdonavan@illingworthrodkin.com)

When quieter pavements are used for purposes of reducing traffic noise levels, one issue of concern is how long will the noise reduction performance be maintained or what is its "acoustic longevity." To address this issue, both the states of California and Arizona have established on-going research projects on a number of different pavement types including the broad groups of asphalt and portland cement concrete (AC and PCC, respectively). These pavements have been monitored routinely with on-board sound intensity measurements since the time of initial construction yielding data over periods of 4–7 years. The pavements consist of 11 PCC pavements of different texture and 13 AC sections of varying constructions. The results of these measurements are presented and discussed. In another approach of evaluating acoustic longevity, pavements of the same construction, but different years of build have also been evaluated to examine longer periods of time. The results of this approach are also presented along with a discussion of the issues involved with the two methods.

3:45

1pNCd3. Quiet concrete pavement surfaces. Robert Rasmussen, Robert Whirledge (The Transtec Group, Inc., 6111 Balcones Dr., Austin, TX 78731, robotto@thetranstecgroup.com), and Paul Wiegand (Natl. Concrete Pavement Technol. Ctr.)

For the last 5 years, a comprehensive evaluation has been conducted on hundreds of concrete pavement surfaces throughout the US, Canada, and

Europe. These data have been synthesized in a manner that allows the highway industry to make important decisions about how best to design and construct concrete pavements to reduce tire-pavement noise levels without compromising other important aspects such as durability, safety, and cost. While a significant amount of variability is present in the data, clear trends have been identified. Some surfaces, such as diamond grinding, appear to be among the best alternatives available today for reducing tire-pavement noise. Meanwhile, recent trials of more unconventional surfaces have yielded mixed results. In the end, the techniques for concrete pavement surfacing today include a vast array of choices. The selection of the best surface for the particular job can now be made more confidently given this improved understanding of the potential impacts due to tire-pavement noise.

4:00

1pNCd4. Noise generation in contraction joints in Portland cement concrete. Tyler Dare (Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031, tdare@purdue.edu) and Robert Bernhard (Notre Dame Univ.)

Contraction joints between slabs of Portland cement concrete have a major impact on measured tire-pavement noise levels. It has been proposed that during the interval when the contact patch of the tire is over the joint, the system acts as a Helmholtz resonator, and that this mechanism is responsible for the majority of the increased noise due to contraction joints. In this paper, this mechanism is investigated through theoretical calculations and experimentation. Intensity probes were used to measure noise near joints of varying widths using Purdue University's Tire-Pavement Test Apparatus (TPTA). The TPTA was used to discover relationships between vehicle speed, joint width, and measured sound pressure and sound intensity spectra near the joint and the tire.

4:15

1pNCd5. On-board sound intensity measurements and results in Florida. Roger Wayson, John MacDonald (US DOT, Volpe Natl. Transportation System Ctr., 3925 Lakeview Dr., Cottonwood Shores, TX 78657, roger.wayson@dot.gov), and Mariano Berrios (Florida Dept. of Transportation)

This paper describes an investigation of on-board sound intensity (OBSI) and concurrent wayside sound levels along multiple roadway surfaces in the state of Florida. By conducting concurrent sampling of the OBSI and wayside noise not only were insights into the pavement texture noise generation at the tire/pavement interface possible but direct comparisons allowed findings on the propagation characteristics as well. Measurement values, pavement rankings, multiple findings, and preliminary statistical modeling will be shown.

MONDAY AFTERNOON, 19 APRIL 2010 GRAND BALLROOM VII/VIII, 1:00 TO 3:00 P.M.

Session 1pNCe

NOISE-CON, Noise, and Structural Acoustics and Vibration: Numerical Methods in Noise Emission and Immission

Bryce K. Gardner, Chair
ESI, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130

Contributed Papers

1:00

1pNCe1. Comparison of sea modeling approaches for vibration transmission through beam structures. Ignatius Vaz, Bryce Gardner (ESI North America, 32605 W 12 Mile Rd., Ste. 350 Farmington Hills, MI 48334-3379, ignatius.vaz@esi-group.com), Vincent Cotoni (ESI Group R&D), and Shuo Wang (NASA Johnson Space Ctr.)

Transmission through beams can become a significant path in many structure-borne problems, especially in aircraft or launch vehicles. Statistical energy analysis (SEA) is widely used for the vibro-acoustic modeling of high-frequency problems. However, in a system-level SEA model, even at higher frequencies, some beams can exhibit low-modal behavior and might not be a good SEA representation. Vibration transmission through beams involves both resonant and non-resonant transmission paths. Typically the non-resonant path is dominant at low frequencies and the resonant path at high frequencies. This paper describes various modeling approaches for describing the transmission through beams over a broad frequency range. SEA modeling techniques will be compared to the hybrid FE-SEA method, energy flow method, and the finite element method for various beam configurations. Modeling beams under pre-stressed conditions is also investigated.

1:15

1pNCe2. Coupling loss factor estimation for plates joined at a point: Analysis and experiment. Mark Ewing, Kranthi Vatti (Aerosp. Eng., Univ. of Kansas, 1530 West 15th St., Lawrence, KS 66045, mewing@ku.edu), and Ignatius Vaz (ESI)

Coupling loss factors for two lightly damped plates joined at a point, but otherwise free to respond in lateral vibration, have been studied both analytically and experimentally. This study forms the basis for on-going estimations of the transient coupling loss factors of the same system. Statistical energy analysis (using VA ONE) and the finite element method (using NASTRAN) have been used to predict the coupling loss factors. The finite element predictions are based on the averaged responses for a range of boundary conditions on the plates, essentially a "pseudo-statistical" approach. Experimentally, two techniques have been used. First, following the approach by Lai and Soom, a force transducer has been used to measure the force between the plates as it is used to attach them. Simultaneously, the kinetic energies in the two plates are measured by three accelerometers attached to each plate during both transient and persistent excitations of the plates, one at a time. The coupling loss factors are estimated by balancing the transmitted energies with the kinetic energies in the two plates. Then, the coupling loss factors were assessed by a more "traditional" approach, namely, balancing the power input to a plate with the energies of the plates. The estimations from these four approaches compare as well as may be expected for these lightly damped plates.

1:30

1pNCe3. Coupling loss factor estimation for plates joined along a line: Analysis and experiment. Mark Ewing, Kranthi Vatti (Univ. of Kansas, 1530 West 15th St., Lawrence, KS 66045, mewing@ku.edu), and Ignatius Vaz (ESI)

Coupling loss factors for two plates joined at a line (a 90-deg bend), but otherwise free to respond in lateral vibration, have been studied both analytically and experimentally. Two cases have been studied: one with no

added damping and one in which one of the plates has a small amount of damping provided by a constrained layer treatment. Statistical energy analysis (using VA ONE) and the finite element method (using NASTRAN) have been used to predict the coupling loss factors. The finite element predictions are based on the averaged responses for a range of boundary conditions on the plates, essentially a "pseudo-statistical" approach. Experimentally, one of the plates was excited while the kinetic energy of the two plates was assessed with a scanning laser vibrometer (scanning a dozen or more points) and four accelerometers on each plate. Both impulse and persistent excitations were used. The coupling loss factors were assessed by balancing the power input with the energies of the plates, also using the estimated modal densities. The estimations from the analytical and experimental approaches compare as well as may be expected for the lightly damped plates case, but very well for the case of one plate having a damping treatment.

1:45

1pNCe4. Development of hybrid one-dimensional finite element/analytical method for analysis of lamb wave propagation characteristics in composit panels. Yong-Joe Kim and Je-Heon Han (Dept. of Mech. Eng., Texas A&M Univ., 315 Eng./Phys. Bldg. Office Wing, College Station, TX 77843-3123, joekim@tamu.edu)

The objective of this study is to develop a hybrid one-dimensional finite element (1-D FE)/analytical method to analyze the Lamb wave propagation characteristics of composite panels. Here, it is proposed to model a composite panel using a finite element approximation in the thickness direction while analytical solutions are assumed in the plane directions. Thus, it makes possible to use small number of finite elements even for high-frequency analysis in computationally efficient manner. By solving the eigenvalue problem of the resulting 1-D FE/analytical matrix equation, the dispersion relations of the Lamb waves propagating in the composite panel can be calculated in wave number/frequency domain. Note that the dispersion relations can be used to determine the frequency, wavelength, and wave speed of a specific Lamb wave mode that can be used to monitor the specific structural defects of the composite panel. For the purpose of validating the proposed method, the dynamic responses of a composite panel are measured. The measured data are transformed into wave number/frequency domain by applying spatial Fourier transform, which result in measured dispersion relations. It is shown that the dispersion relations obtained by using the proposed hybrid method agree well with the measured dispersion relations.

2:00

1pNCe5. Experimental validation of finite element–boundary element model dynamic strain model under diffuse acoustic field loading. Ben Tsoi (Jet Propulsion Lab., 4800 Oak Grove Dr., Pasadena, CA 91109, wan.b.tsoi@jpl.nasa.gov), Bryce Gardner, and Vincent Cotoni (ESI Group NA)

Structural finite element (FE) models naturally output displacement or acceleration response data. However, they can also be used to compute stress, internal forces, and strain response. When coupled with a boundary element model (BEM) of the fluid surrounding the structure, a fully coupled analysis can be performed. Modeling a diffuse acoustic field in the BEM fluid provides an excitation like that found when the structure is placed in a reverberation chamber. Fully coupling the structural FE model to the acoustic BEM model provides a means to predict not only the acceleration response of the panel to diffuse field loading but also the ability to predict the dynamic stress and strain response. This type of model has been available with current predictive tools, but experimental validation of the prediction of dynamic stress or strain is difficult to find. An aluminum panel was instrumented with accelerometers and strain gauges and hung in a reverberation room and subjected to a diffuse acoustic field. This paper presents the comparison of the experimental and predicted results.

2:15

1pNCe6. Modeling the low-, mid-, and high-frequency response of poro-elastic materials in vibro-acoustics applications. Vincent Cotoni (ESI Group, 12555 High Bluff Dr., San Diego, CA 92130, vincent.cotoni@esi-group.com)

A method is presented for representing layered poro-elastic materials and acoustic fluids in a finite element (FE) model of a structure. The method is based on the hybrid FE-statistical energy analysis (SEA) formulation that allows the coupling of FE and SEA descriptions of various subsystems in a model. By making use of the dynamic properties of poro-elastic materials and acoustic fluids (in terms of wavelength and uncertainty), the method is made particularly fast and appropriate for the prediction of the vibro-acoustic response at midfrequencies. The method allows to quickly incorporate acoustics and/or noise control treatment in an existing finite element model of a structure. Given the frequency range of applicability, it complements the existing low-frequency methods (structure and poro-elastic material described with FE, acoustic fluid described with FE, infinite elements, or boundary elements) and high-frequency methods (structure, poro-elastic material, and acoustic fluid described with SEA). The formulation is derived and a number of validation examples are presented.

2:30

1pNCe7. A comparison of visco-thermal models for the vibro-acoustic analysis of hearing aids. Júlio A. Cordioli, Gustavo Martins, and Roberto Jordan (Federal Univ. of Santa Catarina, Campus Univ.-Trindade, Federal Univ. of Santa Catarina Campus Universitrio-Trindade Florianópolis, SC 88090-400 Brazil, cordioli@emc.ufsc.br)

The performance of a hearing aid is directly related to its vibro-acoustics characteristics, which must be taken in account during the design process. Different vibro-acoustic aspects of a hearing aid may be analyzed, including the structural transmission from the loudspeaker to the microphone, which may lead to an unwanted feedback, or the frequency response of the system composed by loudspeaker, tube, earmold and ear channel (considering a behind-the-ear hearing aid). In view of the reduced dimensions of the system it is argued that the vibro-acoustic analysis of such system must take in account visco-thermal effects in the description of the propagation of acoustical waves. In this paper, a simplified system is considered including the loudspeaker membrane, tube, and coupler (element commonly used for testing hearing aids). Results for three different approaches for the vibro-acoustic analysis of the system are compared with experimental data: (i) standard acoustical wave model, (ii) the low reduced frequency model, and (iii) the full linearized Navier–Stokes model. Results show that for the case considered the visco-thermal effects must be taken in account, while the choice between models (ii) and (iii) must be made in view of the system geometry.

2:45

1pNCe8. Systematic approach to passenger vehicle air induction tuning. Antoni Szatkowski, Sean Li, and Jaspal Sandhu (NVH Development & Eng., Chrysler LLC, P.O. Box 21-8004, Auburn Hills, MI 48321-8004, as32@chrysler.com)

Air induction sound level and quality is a key contributor to in-vehicle powertrain NVH performance. This paper describes the process of tuning the air induction systems (AISs) at Chrysler LLC. Simple sound quality metrics are developed to quantify the AIS sound quality. Data acquisition and processing is standardized. A new methodology is presented, how to tune the AIS to meet the desired sound quality, and to optimize the cost, weight, and packaging space. Practical examples are presented to demonstrate how the specific targets are being achieved.

Session 1pNCf

NOISE-CON, Noise, and Signal Processing in Acoustics: Noise Source Localization

Karim Haddad, Cochair
Bruel Kjaer SVM A/S, Skodsborgvej 307, 2850, Naerum, Denmark

Jim Thomspon, Cochair
Bruel & Kjaer North America, 6855 Commerce Blvd., Canton, MI 48187

Contributed Papers

3:30

1pNCf1. Influence of various uncertainty factors on the result of beamforming measurements. Markus Mller-Trapet and Pascal Dietrich (Inst. of Tech. Acoust., RWTH Aachen Univ., Neustraße 50, 52066, Aachen Germany, mmt@akustik.rwth-aachen.de)

In a previous publication [Mller-Trapet and Dietrich (2009)], a virtual measurement environment for sound-source localization on vibrating structures was presented. Based on surface velocity data obtained from laser-scanning-vibrometry measurements, the boundary-element-method is used to simulate the sound radiation from a vibrating plate toward a microphone array under ideal conditions. The advantage of this approach is that the measurement conditions can be perfectly controlled and real sources can be considered, without restrictions on the type of source. The virtual measurement environment will now be used to investigate the effect of some of the uncertainties that can be encountered during beamforming measurements. For the most common planar array geometries, the beamforming source maps will be calculated for varying signal-to-noise ratios and different array imperfections (uncertainties in the location of the microphones and deviation from the omni-directional directivity pattern of the microphones). As a measure of comparison, the two-dimensional normalized cross-correlation between the ideal source map and the source map with added uncertainties will be evaluated and discussed.

3:45

1pNCf2. An analysis of the Doppler shifted aircraft noise signature as measured by a hydrophone located within an evanescent field and implication for localization. David R. Dall'Osto and Peter H. Dahl (Dept. of Mech. Eng. and Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Simultaneous measurements of the acoustic pressure field across the air-water interface, at site in Puget Sound, WA, are analyzed in terms of the Doppler signature of propeller driven aircraft. Measurements in the 50–300-Hz range are considered for various aircraft over flights. The Doppler shift recorded in air by a microphone placed 3 m above the air-water interface is modeled using standard Doppler theory. The hydrophone, placed at a depth 2.5 m below the air-water interface, reveals two Doppler shifts; one associated with the propagating transmitted acoustic field and the other associated with an evanescent field. A parametrized model for these two observed Doppler shifts is presented, utilizing ray theory and Snell's law applied to a plane wave decomposition of the source wave fronts with, $kH \gg 1$, where H is the source height and k is the wave number. It will be shown that analysis of the underwater acoustic field can be used to determine the altitude and closest point of approach (CPA) of a passing tonal source, with support from experimental data recorded by the hydrophone lying within the detection limits of the evanescent field. [Work supported by Washington Sea Grant and ONR.]

4:00

1pNCf3. Interior Fourier near-field acoustical holography using energy density. Zachary A. Collins (Dept. of Mech. Eng., Brigham Young Univ., Provo, UT 84602, zcollins26@gmail.com), Kent L. Gee, Scott D. Sommerfeldt, and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Near-field acoustical holography (NAH) is used to reconstruct three-dimensional acoustic fields from a two-dimensional planar measurement. During previous work at BYU, a method has been developed called energy-based near-field acoustical holography which reduced the number of needed measurements by 75%. Other recent advances have expanded the theory to interior spaces where multiple sources and/or reflections are present. This paper presents a new method for reconstructing interior acoustic parameters using Fourier NAH and a single plane of energy density measurements. Energy density is measured using a six-microphone array. First, the probe measurements are used to create a Hermite surface pressure interpolation on two separate planes. These two planes are used to approximate the normal particle velocity as well as to separate the incoming and outgoing waves using the spatial Fourier-transform method. Once separated, traditional Fourier NAH is used to reconstruct the pressure and normal particle velocity at any point in space. Analytical and experimental results are shown and compared to exterior Fourier NAH approximations. Other drawbacks and benefits are discussed.

4:15

1pNCf4. Conformal mapping based on spherical array measurements: Noise source identification in a car cabin. Karim Haddad (Bruel & Kjaer SVM A/S, Skodsborgvej 307, 2850 Naerum, Denmark, khaddad@bksv.com)

Spherical beamforming based on spherical array is a technique well suited for in-cabin noise source identification, especially from mid- to high frequencies. One aspect of these imaging techniques based on acoustic measurements is how to link detected acoustic sources with the environment. Two main alternatives are currently used. The first possibility is based on pictures of the environment: pictures are superimposed on the acoustic map for an easy identification of sources. The main advantages are its quickness for setup and also it is a good solution to efficiently keep track of any change in the environment. The second possibility is the use of a model describing the objects around the microphone array. It is an interesting solution to describe precisely the medium. But also in terms of accuracy, it provides the correct distances from objects to the acoustic sensor: the level of sources could be more accurately determined. This last option can therefore generate an acoustic map conformal to the surrounding environment. In this paper, we use the conformal mapping technique in a case study concerning tests done in a car cabin using a spherical array.

4:30

1pNCf5. A user friendly, cost-effective and portable tool for locating multiple arbitrary sound sources in three-dimensional space in real time. Na Zhu and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202, linda_zhuna@hotmail.com)

In many engineering applications, it is highly desirable to quickly and easily locate sound sources even if the accuracy and spatial resolution are not very high. Most systems and tools available in the market are based on beamforming technology, which are portable, easy to use, and easy to understand. However, beamforming requires a prior knowledge of the general location of a target source and a large number of microphones, which increases the overall costs. Also, its frequency range is limited by the size of the microphone array. In this paper, a user friendly, cost-effective and portable system for locating multiple arbitrary sound sources in three-dimensional space in real time is presented. In particular, this system requires four microphones only, covers the entire audible frequency range, and does not require any prior knowledge of whereabouts of the target sources. The underlying principle of this methodology is a hybrid approach consisting of modeling of sound radiation from point sources in a free field, triangulation, and signal processing techniques. Both numerical and experimental validations are demonstrated and an error analysis in locating arbitrary sources is performed. The main cause of errors is found to be due to the fact that the basic formulations are for a free field, whereas in practice the measurement environment is a non-free field.

4:45

1pNCf6. Panel contribution analysis for a large complex structure. Sandeep Mylavarapu and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., Petroit, MI dz2727@wayne.edu)

This paper presents a newly developed panel contribution analysis using the Helmholtz equation least squares method for a large complex structure. The approach requires only one set of the acoustic pressures measured in the near field of a vibrating structure. Using these input data, the normal surface velocity and acoustic pressure are reconstructed, from which the normal acoustic intensity distribution is calculated. Next, the normal acoustic intensity is correlated to the acoustic pressure at any field point of interest. In this way, the relative contributions from individual panels toward acoustic radiation at any field point can be ranked, and the critical panel(s) of a vibrating structure can be identified. Numerical simulations using this approach for identifying critical panels inside a passenger compartment are presented.

MONDAY AFTERNOON, 19 APRIL 2010 LAUREL A/B, 1:20 TO 5:00 P.M.

Session 1pNSa

Noise, Physical Acoustics, and INCE: Rocket Noise Environments II

Kent L. Gee, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

R. Jeremy Kenny, Cochair
NASA Marshall Space Flight Center, Huntsville, AL 35812

Invited Papers

1:20

1pNSa1. Directivity indices for rocket noise modeling: Measurement considerations. Jarom H. Giraud and Kent L. Gee (Dept. of Phys. and Astronomy, Eyring Sci. Ctr., Brigham Young Univ., 84602 Provo, UT, jaromgiraud@gmail.com)

Long-standing rocket noise models use far-field directivity indices to predict acoustic radiation loading on and near a launch vehicle. This approach raises a number of critical questions including where does the geometric far field begin for the frequencies of interest? and how does probable nonlinear propagation of noise away from the source affect these indices? Data collected on a static, horizontally fired GEM-60 solid rocket motor provide some insight into these questions. In the near field, microphones were located along a line approximately eight nozzle diameters from the shear layer. In addition, microphones were placed along two radials near the peak directivity angle at 63, 125, and 250 nozzle diameters. Analysis of these data indicates that nonlinear propagation dramatically affects the high-frequency directivity patterns along the peak radiation angles. In addition, the extended length of the source likely pushes the onset of the geometric far field to beyond 100 nozzle diameters for low frequencies.

1:40

1pNSa2. Acoustic measurement in the static firing tests of solid rocket motors. Tatsuya Ishii, Hideshi Oinuma, Junichi Kazawa (Aviation Program Group, JAXA, 7-44-1 Jindaiji-Higashi-Machi, Chofu, Tokyo 182-8522, Japan), Kota Fukuda, Seiji Tsutsumi (JEDI Ctr., JAXA, 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan), Kyoichi Ui (Space Transportation Mission Directorate, JAXA, Tsukuba-shi, Ibaraki 305-8505, Japan), and Kenji Minesugi (Inst. of Space and Astronautical Sci., JAXA, Sagamihara, Kanagawa 229-8510, Japan)

Acoustic management is essential for establishing a reliable and cost-competitive rocket system. A solid propellant motor during its lift-off imposes high levels of pressure and vibration on the fairing of the rocket, sometimes leading to crucial damage of the payload. One approach to resolve this problem is to decrease the sound sources caused by high-speed plume of the rocket booster. Japan Aerospace Exploration Agency (JAXA), aiming at the advanced solid rocket, started computational prediction that helps design a launch-pad

with less acoustic impact. For validating the computation codes and modeling the acoustic characteristics, the experimental acoustic data of solid rocket motors have long been desired. Fortunately, we had opportunities of the ground firing tests with several motors. In this presentation, the acoustic measurement including the sensors, the set-up, and the data reduction carried out in these tests will be discussed.

2:00

1pNSa3. A model for the prediction of community noise from launch vehicles. Kenneth J. Plotkin (Wyle Labs., 241 18th St. S., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wyle.com)

A model, RNOISE, has been developed for prediction of far-field community noise from the launch and boost phase of launch vehicles. It follows the moving source principles of an early NASA model [Wilhold *et al.*, NASA Report No. TN D-1832 (1963)] but uses a noise source and directivity model developed by Sutherland [AIAA Paper No. 93-4383 (1993)] together with modern propagation algorithms as employed in the time simulation aircraft models NMSIM (NoiseMap Simulation) and AAM (Advanced Acoustic Model). RNOISE, first used for environmental analysis in the mid-1990s, is a spectral time simulation model, generating predictions of one-third octave band spectra on the ground as a function of time. Predictions may be made at a single point or on a grid. Recent improvements to RNOISE are the ability to read flight path data from a PCBoom format trajectory file, and output of time history grids in the TIG format used by NMSIM, AAM, and RNM (Rotorcraft Noise Model). This permits presentation of results as animations as well as traditional static contours.

2:20

1pNSa4. Rocket motor microphone investigation. Eric Herrera, Debbie Pilkey (ATK Launch Systems, P.O. Box 707, Brigham City, UT 84302), Kent L. Gee, Jarom H. Giraud, and Devin J. Young (Brigham Young Univ., Provo, UT 84602)

Identification of optimal microphone configurations and instrumentation is underway as a continuous process improvement effort at ATK Space Systems, where large solid rocket motors (SRMs) have been tested for decades. Rocket motor tests provide a unique environment which cannot be fully duplicated in a laboratory. This is especially true for the largest SRM ever to be tested, Demonstration Motor I (DM-I), which is the five-segment first stage of the NASA Constellation program's Ares I launch vehicle. During the DM-I firing, two separate arrays of pressure transducers with various vent tube configurations, microphones, accelerometers, and temperature sensors were placed at locations near the motor. The objectives were to better determine the vibration levels and temperature to which acoustic gages are subjected and to examine their comparative performance for different vent tube configurations, particularly at low frequencies (10–30 Hz). Results from the DM-I and corroborating laboratory tests reveal very different low-frequency responses for the pressure transducers that depend critically on vent tube configuration. This and other issues are discussed.

2:40

1pNSa5. Condenser microphone protective grid correction for high-frequency measurements. Erik N. Lee and Reginald M. Bennett (Marshall Space Flight Ctr., MS ER42, Huntsville, AL 35812)

Use of a protective grid on small diameter microphones can prolong the lifetime of the unit, but the high-frequency effects can complicate data interpretation. Analytical methods have been developed to correct for the grid effect at high frequencies. Specifically, the analysis pertains to quantifying the microphone protective grid response characteristics in the acoustic near field of a rocket plume noise source. A frequency response function computation using two microphones will be explained. Experimental and instrumentation setup details will be provided. The resulting frequency response function for a B&K 4944 condenser microphone protective grid will be presented, along with associated uncertainties.

3:00—3:20 Break

3:20—5:00 Panel Discussion

Session 1pNSb

Noise and Architectural Acoustics: Soundscape Concert II

Alex U. Case, Cochair
Fermata Audio & Acoustics, P.O. Box 1161, Portsmouth, NH 03802-1161

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Einsteinufer 25, Secr TA 7, Berlin, 10587, Germany

Invited Papers

1:00

1pNSb1. Assessing the acoustical attributes of preserved historic structures. Eric L. Reuter (Reuter Assoc., LLC, P.O. Box 4623, Portsmouth, NH 03802, ereuter@reuterassociates.com)

Acoustical attributes of several preserved historic structures have been measured using contemporary techniques and metrics. These include outdoor-to-indoor and indoor-to-indoor impulse response and noise reduction measurements. Combined, these measurements allow for simulation, both quantitative and qualitative, of the indoor soundscapes of the past. These data will be presented, along with audio examples of historical noise sources, as experienced from indoors.

1:40

1pNSb2. Listening to soundscape as the music of life. Caroline Cance (INCAS3, Dr. Nassaulaan 9, P.O. Box 797, Assen 9400 AT, The Netherlands, ccance@gmail.com), Danièle Dubois and Hugues Genevois (CNRS-Univ. Paris 6-MCC, Paris 75015, France)

The musician M. Shafer introduced, in 1970, the word and the concept of soundscapes as "the tuning of the world," driving our attention to the quality of the sounding environment, therefore comparable to music. The concept of soundscape has been developed within acousticians' communities, and the research slowly evolved from "noise effect," "noise control," or "noise annoyance" to "sound quality" or "soundscape" and "soundscaping." Also, around 70 electroacoustics music has promoted every day noises as musical sounds, and nowadays computer music blurs the boundaries between musical instruments and new technological devices and tools. We would like to present here our research conducted in perception and evaluation both in music (and more precisely on digital music) and in everyday noise and soundscapes to discuss some of the following questions: What makes a noise a musical sound? Its acoustic properties? That it is produced by a specific (musical) instrument? The time, places (streets? concert halls?) or ways one hears/listens to it (personal involvement, purpose)? Are the boundaries between music and noise the same in different communities (acousticians, musicians, instrument makers, every one in the street)? How do the concepts of music/noise evolve in time and space, and people?

2:20

1pNSb3. The daily rhythm of the soundscape "Nauener Platz" in Berlin. Brigitte Schulte-Fortkamp (Inst. of Fluid Mech. and Eng. Acoust., Technische Universitt Berlin, 10587 Berlin, Germany)

Rhythm in soundscape characterizes the time structure of the soundscape with its distinctive sound events embedded in comprehensive acoustical scenery. The "Nauener Platz" evaluation was balancing between acoustic measurements, architectural planning, and the expertise from people living in the area to contribute to the reconstruction of the place. Information of a new understanding and concept of a public place was underlying the development of the new Soundscape based on noise reduction and audio islands playing the sound people like to live with in that area. This Soundscape will be provided here in musical performance also explaining the cultural and social issues behind. The Soundscape Project is a module of the Project "Nauener Platz-Remodelling for Young and Old" in the framework of the research program "Experimental Housing and Urban Development (ExWoSt)" of the "Federal Ministry of Transport, Building, and Urban Affairs (BMVBS)" by the "Federal Office for Building and Regional Planning (BBR)." It is related to the fields of research (ExWoSt) concerned with "Innovation of Urban Neighbourhoods for Families and the Elderly."

3:00—3:20 Break

3:20

1pNSb4. Measurement of whole hike soundscapes in a national park setting. G. (Randy) Stanley (Natural Sounds Program, U.S. Natl. Park Service, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

Measurements of whole hike soundscapes were made in Mount Rushmore National Park. The goal was to capture the time and space varying aural stimulus that a hiker experiences while walking through an environment. Two elements were employed: A portable binaural headset recording system and an array of audio recorders arranged at intervals along the trail. Various elements of the recorded soundscapes will be summarized and reproduced in part for the audience. In addition, expected analyses and anticipated challenges will be discussed.

4:00

1pNSb5. Exploring the soundscape of nature. Gary W. Siebein, Keely M. Siebein, Lucky Tsaih, Sang Bum Park (Univ. of Florida School of Architecture, P.O. Box 115702, Gainesville, FL 32611), Chris P. Jones, Reece Skelton, and Joshua Fisher (Siebein Assoc., Inc., Gainesville, FL 32607)

A study of a state park situated along a spring fed river was conducted to try to understand the nature of the soundscape of a park intended to be used by people for recreational purposes as well as to preserve the qualities of the natural ecosystems in the area. The study explores the qualitative and quantitative aspects of the ambient sound as composed of a series of specific acoustic events related to natural sounds of wildlife, water flows and weather, human use of the park, maintenance operations by park staff, and off-site noise intrusions from adjoining land uses. The acoustic "rooms" of the site are identified by the characteristics of their impulse responses. The rooms are linked by acoustic itineraries related to each of the user groups. Modeling of animal, weather, and human sounds is used to construct the acoustic rhythms of the site from limited, practical data acquisition of existing sounds. Recordings of real and simulated sound fields from the park were evaluated by listeners to determine qualitative and quantitative relationships of real and modeled sounds.

4:40

1pNSb6. Ocean dreams. Bennett M. Brooks (Brooks Acoust. Corp., 30 Lafayette Square, Vernon, CT 06066, bbrooks@brooksacoustics.com)

We arise by the sea, the genesis of life. We move inland via various forms of transportation, observing daily activities, ending our visit full circle with the solice at water's edge.

MONDAY AFTERNOON, 19 APRIL 2010 HARBORSIDE B, 1:30 TO 5:00 P.M.

Session 1pNSc

Noise and INCE: Noise and Its Control in Complex and Urban Environments

Kai Ming Li, Cochair
Purdue Univ., Dept. of Mechanical Eng., 140 Martin Jischke Dr., West Lafayette, IN 47906

Siu-Kit Lau, Cochair
Univ. of Nebraska-Lincoln, School of Architectural Eng., 1110 S. 67th St., Omaha, NE 68182-0681

Invited Papers

1:30

1pNSc1. Urban acoustic and seismic noise measurements in Baltimore. Donald G. Albert and Stephen N. Decato (ERDC Cold Regions Res. and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755)

Acoustic and seismic sensors have good potential for passive, non-line-of-sight, detection, localization, and monitoring of natural and human activities. However, to fulfill this potential, sensor algorithms must be designed to not only robustly identify signals of interest but also to cope with environmental effects including high levels of background noise often found in urban areas. This paper presents experimental measurements conducted in Baltimore and at other locations to investigate noise levels in rural, urban, and industrial environments for audio frequencies between 6 Hz and 20 kHz and seismic frequencies from 6 to 1000 Hz. Because of the interest in large-scale Earth studies, there are hardly any previous seismic measurements in this high-frequency band and also most were done in remote areas specifically to avoid cultural noise. The measured urban seismic spectra tend to have more structure and variation in shape from one location to another compared to the acoustic spectra, and a more rapid decay with increasing frequency, usually about 30 dB from 30 to 600 Hz. The urban acoustic peak frequency is usually around 60 Hz, while the peak seismic frequencies are usually less than 25 Hz. [Work funded by U.S. Army.]

1:50

1pNSc2. Investigation of noise barriers with resonators by the indirect boundary element method. Markus Auerbach (Federal Hwy. Res. Inst. (BASt), Bruederstr. 53, D-51427 Bergisch Gladbach, Germany, m.auerbach@bast.de), Andreas Bockstedte, Olgierd Zaleski, Marian Markiewicz (Novicos GmbH, D-21073 Hamburg, Germany), Otto von Estorff (Hamburg Univ. of Technol. (TUHH), 21073 Hamburg, Germany), and Wolfram Bartolomaeus (Federal Hwy. Res. Inst. (BASt), D-51427 Bergisch Gladbach, Germany)

Noise barriers are the most common solutions to protect against noise pollution. They are very effective in the mitigation of road noise, but their high effectiveness often involves undesirably huge constructions. Following the demand for improving the efficiency without increasing the heights, a new procedure has been developed, which allows for extending a typical noise barrier by a line of resonators in a quite efficient way. It is based on numerical investigations including the indirect boundary element method. Several

representative calculations are shown which demonstrate how the new approach can be used. The procedure turns out to be computationally very powerful and it seems to be a very promising step toward a more efficient design of sound barriers featuring high effectiveness at moderate overall heights. Measurements at a 20 m wide and 4 m high noise barrier with and without the device will be presented.

2:10

1pNSc3. Modeling of sound fields in urban streets using a one-dimensional transport equation model. Yun Jing and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

This paper presents a one-dimensional transport equation model for acoustic field predictions in urban streets. Considering urban streets as elongated enclosures, acoustic properties of the surfaces are typically quite different. For example, the top and two ends are often open and the streets are usually reflective. A recently developed one-dimensional transport equation [Jing *et al.*, J. Acoust. Soc. Am. **125**, 2735 (2009)] is extended to model non-uniform absorption and scattering coefficients and is applied to urban streets acoustic predictions. This paper discusses preliminary experimental results obtained from a tenth scale-model to validate the model.

2:30

1pNSc4. Numerical and experimental studies for reducing the transmission of low frequency sound through windows. Andrew M. Jessop, Kai Ming Li, and J. Stuart Bolton (Ray W. Herrick Labs., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031)

Windows are a common path for low-frequency sound transmission from outside to inside due to their light weight and low damping. The use of non-periodic stiffening elements to increase window transmission loss without reducing visibility is considered here. The effect of elastic mounting of the window to the surrounding structure as a possible contributor to low-frequency noise transmission was also examined. A finite element model allowing for panel stiffening, either by clamping or by adding variable-stiffness beams within the panel, was constructed; the model allowed for variable positioning of such elements. Several double-panel configurations, representing double-pane windows typically used in energy-efficient constructions, were also modeled with various clamped-stiffening elements. Scale-model measurements of transmission loss of aluminum panels were performed to verify the character of the predictions with variable edge stiffness as used in the numerical models. The computational results show that low-frequency sound below 150 Hz can be reduced by several dB through careful use of stiffening elements and that resilient mounting can reduce the transmission loss of low-frequency sound through windows.

2:50

1pNSc5. Parametric study of the sound insulation performance of a ventilation window. S. K. Tang (Dept. of Bldg. Services Eng., The Hong Kong Polytechnic Univ.), D Suen, and S. K. Chan (Campus Development Office, The Hong Kong Polytechnic Univ.)

The sound insulation performance of a specifically designed facade device, which allows natural ventilation but can provide relatively high-sound insulation, is studied experimentally in the present study. The experiment was carried out in two coupled rooms with one of them very reverberant and the other made semi-anechoic. The former acted as the receiver room and the latter contained a linear loudspeaker array (25 loudspeakers) as the noise source. The full scale facade device was installed on the concrete brick wall between the coupled rooms. Nine microphones were used to measure the average sound intensity inside the reverberant room. The effect of the size of the device opening (for ventilation) on the sound insertion loss was then determined by the difference between the average intensities with and without the device (an opened window). The present results show that the present facade device can offer additional acoustical protection of ~12–13 dB(A) compared to that resulted from the conventional openable window design when exposed to traffic noise. The introduction of sound absorption materials at the top internal surface of the window cavity in general can result in a maximum of 2 dB extra benefit.

3:10—3:30 Break

Contributed Papers

3:30

1pNSc6. Three-dimensional pseudospectral time-domain calculations of sound propagation to closed urban courtyards. Maarten Hornikx (Dept. of Mech. Eng., K.U. Leuven, Celestijnenlaan 300b-bus 2420, B-3001 Heverlee, Belgium, maarten.hornikx@mech.kuleuven.be) and Jens Forssén (Chalmers Univ. Tech., Gothenburg S-41296, Sweden)

Because contemporary urban environments suffer from excessive levels of road traffic noise, access to courtyards with low-noise levels is essential in order to offer urban sound environments of high quality with regard to health and perceived sound. Due to limitations in computational resources, accurate prediction of road traffic noise to urban courtyards has mainly been restricted to modeling the courtyards in two dimensions. To calculate sound propagation from road traffic to three-dimensional courtyards, the extended Fourier pseudospectral time-domain method is used here. This numerical domain discretization method solves the linearized Euler equations, where spatial derivatives are evaluated based on an eigenfunction expansion and thereby requires only two spatial points per wavelength. The time iteration is done using a low-storage optimized six-stage Runge–Kutta method. Calculation results, in time and frequency domains, display the difference of treating courtyards in two and three dimensions. As one of the results, it is concluded that the effect of noise abatement in the form of facade absorption treatments is larger for a setting in three dimensions than in two dimensions.

3:45

1pNSc7. Influence of impedance distribution and optimal impedance design in the context of noise barriers. Alexandre Jolibois, Victor W. Sparrow (Grad. Program in Acoust., Penn State, 201 Appl. Sci. Bldg., University Park, PA 16802, auj145@psu.edu), Denis Duhamel (Ecole des Ponts ParisTech, 77455 Marne La Vallée, France), and Jérôme Defrance (Ctr. Scientifique et Technique du Bâtiment, 38400 Saint-Martin-d'Hères, France)

Noise barriers are commonly used to protect people from railways or roads traffic noise. However, the optimal design of those barriers is still an open problem since lots of different results have been obtained, depending on the method utilized. In this paper, instead of focusing on the shape of the barrier itself, a new method involving optimization of the impedance distribution on a control surface of simple geometry and containing the barrier has been used. First, numerical simulations have been carried out with the

two dimensional boundary element method in order to predict the sound pressure level around any impedance scatterer. Analytical solution for the scattering of a line source by an infinite constant impedance circular cylinder has been derived and used for validation. Second the influence of the impedance distribution in the case of a circular cylindrical control surface has also been studied to determine an optimal impedance distribution to control the noise level in a zone close to the barrier. Specific attention has been paid to the scattering of multiple line sources in order to model high-speed train noise sources more accurately.

4:00

1pNSc8. Model of a benefit/cost ratio analysis for comparison of environmental noise barriers. Antonio P. Carvalho and Pedro D. Oliveira (Dept. Civil Eng., Lab. of Acoust., College of Eng., Univ. of Porto, 4200-465 Porto, Portugal, carvalho@fe.up.pt)

The choice for an economically ideal solution of environmental noise barrier must acknowledge both the cost of its main components and the benefits it can provide through time. To achieve a systematic analysis tool an algorithm based on benefit/cost ratio analysis was created. It calculates the benefit/cost ratio for any potential noise barrier. The cost of a barrier can be described with known or quantifiable parameters such as barrier height, thickness, angle, material and initial investment costs, life-cycle costs as maintenance costs, replacement costs due to minor/major accidents, vandalism or defacement, etc. The benefits associated with a solution are defined by computable parameters such as sound absorption index (DLalfa) and airborne sound insulation index (DLr) defined in EN 1793, insertion loss (predicted or measured), and even intangible parameters such as visual intrusion (color and aesthetics) and environmental impact (sustainability). Each benefit is weighed regarding its importance. Using the necessary parameters it is possible to calculate the benefit/cost ratio for any number of years of a life expectancy. An analysis concerning several possible solutions with different life expectancies can therefore be conducted. A general example of application is also presented.

4:15

1pNSc9. Noise attenuation by two-dimensional sonic crystals made of rubber crumb. Jose Sanchez-Dehesa, Victor M. Garcia-Chocano, Daniel Torrent, Francisco Cervera, Suitberto Cabrera (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de vera s.n., ES-46022, Spain, jsdehesa@upvnet.upv.es), and Francisco Simon (Instituto de Acustica, ES-28006 Madrid, Spain)

It is shown that sonic crystals consisting of cylindrical rods made of rubber crumb can be used as efficient structures to attenuate noise in a wide range of frequencies. A structure made of three rows of cylinders (3-m height) has been fabricated and its isolation properties have been characterized in a transmission chamber. Two different numerical algorithms have used to simulate the experimental characterization. A good agreement between theory and experiment is obtained and gives support to design barriers based on this technology. [Work supported by MICIIN of Spain.]

4:30

1pNSc10. Annoyance response to mapped noise from combined road-traffic and railway system. Jin-Hee Son, Tae-Ho Park, Joon-Hee Ko, Seo Il Chang (Dept. of Environ. Eng., Univ. of Seoul, Seoul 130-743, Republic of Korea), and Kun Lee (Univ. of Seoul, Seoul 130-743, Republic of Korea)

To study annoyance response to combined transportation noise from road-traffic and railway, a social survey has been performed in a highly populated residential area in the city of Seoul, South Korea. To determine the noise exposure level in an interviewed household, a noise mapping software package was utilized and the noise exposure level from each single source was calculated separately and combined to get total exposure level. The prediction was partially validated by measurements. The households in the study area are classified into acoustically differentiated groups which are named as the road-traffic noise dominant, railway noise dominant, and combined noise dominant, respectively. %HA (percentage of highly annoyed people) curves to noise exposure level are presented for the acoustically classified three groups. Comparison of the %HA curves shows that more people are highly annoyed to railway noise than road-traffic noise, but the curve for combined noise is much closer to the %HA curve for road-traffic noise rather than the %HA curve for railway noise. Multiple-regression and path analysis of the survey data give the direct and indirect quantitative relationships between annoyance and noise and non-noise factors. Personal sensitivity to noise and satisfaction of living environment are primary contributing factors.

4:45

1pNSc11. Noise mechanisms in motorcycle helmet noise. M. J. Carley (Dept. of Mech. Eng., Univ. of Bath, Bath BA2 7AY, United Kingdom, m.j.carley@bath.ac.uk), N. J. Holt (Bath Spa Univ., Bath BA2 9BN, United Kingdom), and I. Walker (Univ. of Bath, Bath, BA2 7AY, United Kingdom)

A unique set of results on the acoustics of motorcycle helmets has been gathered during road tests on a rider wearing a representative modern helmet. The data were collected during a study of the noise which can cause hearing damage and, possibly, distraction in riders. They consisted of simultaneous measurements of noise at the rider's ear and unsteady pressure on the helmet surface, combined with GPS measurements of rider position and speed. These signals have been analyzed to reduce the coherent structures in the turbulent flow responsible for noise generation. The identified structures appear to be produced by a vortex street shed by the motorcycle windscreen. The internal and external pressures proved to be poorly correlated over most of the frequency range, which has been identified as a result of the insertion loss of the helmet. The implications of these findings are that the majority of the variation in helmet noise is a function of such extrinsic factors as motorcycle configuration and rider build and posture. Efforts to reduce the harmful effects of noise in motorcycling should, then, move to studying the whole system of rider, helmet, motorcycle, and external environment.

Session 1pNSd

Noise, ASA Committee on Standards, and INCE: Wind Turbine Noise

Robert D. Hellweg, Jr., Cochair
EPSILON Assoc., 3 Clock Tower Pl., Ste. 250, Maynard, MA 01754

Kenneth H. Kaliski, Cochair
Resource Systems Group, 55 Railroad Row, White River Junction, VT 05001-7130

Chair's Introduction—1:15

Invited Papers

1:20

1pNSd1. Calculating annualized sound levels for a wind farm. Kenneth Kaliski and Eddie Duncan (Resource Systems Group, 55 Railroad Row, White River Junction, VT 05001, kikaliski@rsginc.com)

Modeling done for wind farms usually focuses on calculating a worst-case short term average sound level. However, the impact to homes is not simply defined by a single meteorological condition. Rather, a more complete picture of the impacts is given by calculating sound levels under various meteorological conditions that occur during the year. The actual sound level at a receiver will depend on variations in atmospheric stability, wind speed, wind direction, and other parameters that change hourly. This paper will describe a method to calculate hourly sound pressure levels for individual receivers over the course of an 8760 h year and give examples of different wind farm configurations and how they affect annualized sound levels.

1:40

1pNSd2. Low frequency sound and infrasound from wind turbines: A status update. Robert D. O'Neal, Robert D. Hellweg, Jr., and Richard M. Lampeter (Epsilon Assoc., Inc., 3 Clock Tower Pl., Ste. 250, Maynard, MA 01754, roneal@epsilonassociates.com)

A common issue raised with wind energy developers and operators of utility-scale wind turbines is whether the operation of their wind turbines may create unacceptable levels of low frequency noise and infrasound. In order to answer this question, one of the major wind energy developers commissioned a scientific study of their wind turbine fleet. The study consisted of three parts: (1) a world-wide literature search to determine unbiased guidelines and standards used to evaluate low frequency sound and infrasound, (2) a field study to measure wind turbine noise outside and within nearby residences, and (3) to compare the field results to the guidelines and standards. The guidelines and standards evaluated were audibility including infrasound; ANSI S12.2 for interior sounds—both acceptability of low-frequency sounds in bedrooms, schools, and hospitals and perceptible rattles and vibration; ANSI S12.9 Part 4 for thresholds of annoyance and beginning of rattles; and certain European criteria for low frequency and infrasound. This paper presents the results of the detailed study and concludes that there should be no adverse public health effects from infrasound or low-frequency noise at distances greater than 1000 ft from the wind turbine types measured.

2:00

1pNSd3. Potential noise reduction from mast trailing vanes as downwind-design wind turbine retrofits. Mark Storm (URS Corp., 1615 Murray Canyon Rd., Ste. 1000, San Diego, CA 92108, mark_storm@urscorp.com)

Older wind turbine generators (WTGs) featuring rotors that are "downwind" of the nacelle mast structure have the reputation of being noisier than their "upwind" design counterparts on the argument that the aerodynamic wake from the mast introduces noise-producing turbulence at the interaction area of the passing rotor blades. Possible methods of improving the wake, for the purpose of reducing this blade-interaction turbulence and hence noise production, could include the addition of a flat plate or similar structure or device to the trailing side of a mast structural element. Realization of such a conceptual device would likely be a retrofit to a pre-existing downwind-design WTG—if mast-to-rotor clearance and other practical parameters allow. This study attempts to quantify the noise reduction potential of a sample retrofit concept, using an estimation technique similar to what one might find in the realm of commercial HVAC technology, where axial fan blade interaction with wakes from stators, vanes, pipes, grilles, or guards is a common noise-producing concern.

Session 1pSP

Signal Processing in Acoustics, Engineering Acoustics, and Physical Acoustics: Battlespace Acoustics

Michael V. Scanlon, Chair
U.S. Army Research Lab., 2800 Powder Mill Rd., Adelphi, MD 20783

Chair's Introduction—1:00

Invited Papers

1:05

1pSP1. Acoustic sensing for urban battlefield applications. Tien Pham (US Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, tien.pham1@us.army.mil)

For years, acoustic sensor systems have been used effectively in a variety of battlefield applications due to its low-cost, low-power, non-line-of-sight and 360 deg coverage capabilities. Acoustic (only) systems can autonomously detect, track, and classify a variety of targets: personnel, ground vehicles, airborne targets, and transient events. In recent years, acoustics has become an important sensing technology due to its complementary sensing capability within multi-modal sensor systems, and its wide-area coverage capability to alert and point other hi-resolution sensors within a network of sensor systems for intelligent-surveillance-reconnaissance (ISR). For example, in a multi-modal sensor system, acoustic can provide enhanced detection and discrimination of targets and activities, and cue imaging sensors to capture the events of interest. Similarly, in a network of ISR sensor systems, acoustic can reduce the search space for other ISR sensors to autonomously track and locate events of interest in real time. In this discussion, we will present an overview of battlefield acoustic applications that are of current interest to military operations: (i) detection and localization of transients, (ii) sensing from mobile and aerial platforms, and (iii) sensing within a network of ISR sensor systems. We will then discuss the signal processing strategies and challenges associated with the applications.

1:25

1pSP2. Effects of atmospheric scattering and refraction on the performance of acoustic direction-finding arrays. D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03768-1290, d.keith.wilson@usace.army.mil), Sandra L. Collier (U.S. Army Res. Lab., Adelphi, MD 20783), and Vladimir E. Ostashev (NOAA Earth System Res. Lab., Boulder, CO 80303)

Refraction, turbulent scattering, and other atmospheric propagation effects complicate the performance of outdoor microphone arrays used to infer source bearings. In principle, with good knowledge of the atmospheric profiles, one can compensate for the refraction effects. However, the random angle-of-arrival variations induced by atmospheric turbulence limit array performance even when the signal-to-noise ratio is high. The turbulent scattering effect can be viewed as a coherence loss between elements of the beamforming array. Most research to date on acoustic signal coherence in the atmosphere has dealt with line-of-sight (LOS) propagation paths. A number of complications characteristic of real turbulence have been introduced; these include statistical inhomogeneity, anisotropy, and intermittency of the turbulent eddies. The LOS theory has also been extended to include sensor displacements longitudinal to the propagating wavefronts as well as transverse displacements. Formulations for the Cramer–Rao lower bound (CRLB) on the angle-of-arrival accuracy have been developed from the LOS theory. Based on limited experimental data and modeling, actual performance appears to be substantially worse than the CRLB. Recently, numerical methods have been introduced for solving second-moment parabolic equations, which offer the possibility of incorporating non-LOS effects into coherence calculations.

1:45

1pSP3. Associating acoustic transients for localization by multiple arrays. Tung-Duong Tran-Luu (2800 Powder Mill Rd., Adelphi, MD 20783, dtranluu@arl.army.mil)

This paper first presents two methods to estimate the location of an acoustic transient based on measurements of its time-of-arrival and direction-of-arrival at a set of microphone arrays. One is a least-square (LS) estimator, the other a maximum likelihood (ML) estimator. Both are compared to the Cramer–Rao lower bound for reference. They are then extended to associate detections (of the same event) at different arrays by a maximum *a posteriori* (MAP) formulation that can sort out outliers and multipath detections while reusing the likelihood values computed in the first part. Simulations show that the LS estimator performs slightly better than the ML estimator when the observation noise does not match the noise model. Both methods exhibit a bias in the range estimate, which accounts for most of the square error. The MAP estimator, applied to live fire data, was accurate and successfully resolved simultaneous multiple targets from outlier and multipath noise.

2:05

1pSP4. Comparison of noise levels measured on elevated acoustic sensor arrays mounted on various aerostats. Christian G. Reiff (2800 Powder Mill Rd., Adelphi, MD 20783)

The army has fielded acoustic systems for the detection of mortar, rocket, and small arms and it has been demonstrated that elevating the acoustic array by mounting the array on an aerostat can improve signal-to-noise levels. The elevated array receives a cleaner, higher amplitude signal while the noise level also increases. The performance of aerostat mounted acoustic sensor arrays for source detection and localization is limited by these increased noise levels caused by higher wind speeds and airship components. These shortcomings can be mitigated somewhat by using microphone windscreens, array location, and signal processing. Data from experiments on a large tethered rigid airship, a smaller tethered non-rigid aerostat, and a medium rigid motorized airship are compared for noise environment and effects on acoustic detection and source localization.

2:20

1pSP5. Refraction corrections in source localization with an elevated acoustic sensor array in the 2007 Yuma experiment. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., 325 Broadway, Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003, vladimir.ostashev@noaa.gov), Christian G. Reiff, Michael V. Scanlon, Sandra L. Collier (U.S. Army Res. Lab., Adelphi, MD 20783), and D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

Performance of ground based acoustic sensor arrays for source detection is limited by several factors, including sound reflections by obstacles, irregular terrain, multipath sound propagation, absorption of sound waves in the ground, and formation of an acoustic shadow zone in the upwind direction. These shortcomings can be overcome by suspending acoustic sensor arrays below tethered aerostats. A comprehensive experiment in source localization was carried out in October of 2007 near Yuma, AZ. A sensor array was suspended below an aerostat, and ray refraction corrections were calculated for the apparent source coordinates. Significant bias errors in the source coordinates were found if a straight-line approximation for sound propagation in the atmosphere is used. However, by taking into account refraction of sound signals due to vertical profiles of temperature and wind velocity, it is possible to significantly reduce the localization errors even with a rough approximation of these meteorological profiles. If actual profiles were used in refraction corrections, localization errors would probably be further reduced.

2:35

1pSP6. Comparison of the performance of acoustic beamformers under differing atmospheric conditions. Sandra L. Collier, Leng Sim, Duong Tran-Luu (U.S. Army Res. Lab., RDRL-CIE-S, 2800 Powder Mill Rd., Adelphi, MD 20783, scollier@arl.army.mil), D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03768), and Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305)

It is well known that atmospheric turbulence can negatively impact the performance of acoustic beamformers. While many beamformers, especially adaptive ones such as minimum variance distortionless response (MVDR), may be robust when the turbulent fluctuations are mild to moderate, they fail when the fluctuations are large. Other methods, such as maximum likelihood estimation, may be used to mitigate the effects of turbulence by directly incorporating the physics of the propagation medium into the assumed model of the acoustic signal (through the covariance and mean). When using synthesized data, the previously developed maximum likelihood estimator (MLE) for the azimuthal angle of arrival was found to outperform classical beamformers. However, in reality the atmosphere does not exhibit the exact behavior of the assumed atmospheric model, or all the required input parameters, such as the meteorological data, for the atmospheric model are not available. Therefore, we compare the performance of methods such as the MLE to that of classical and parametric methods, such as MVDR, multiple signal classification, and matched subspace detector, for data collected during a variety of atmospheric conditions. We critically examine the expense of gained accuracy over computational speed.

2:50

1pSP7. Impact of a collection of buildings on line-of-bearing estimates. John M. Noble, W. C. Kirkpatrick Alberts, II, and Mark A. Coleman (U.S. Army Res. Lab., ATTN: RDRL-CIE-S, 2800 Powder Mill Rd., Adelphi, MD 20783)

The urban environment presents a very complex environment for the propagation of acoustic signals. To study some of these effects, a series of field experiments was conducted looking at propagation around urban structures. The third experiment in the series involved propagation around a small collection of urban structures. As part of this experiment, a series of triangular arrays was scattered around the area to examine the effect of urban structures on the line-of-bearing calculation. This presentation will present the results of these measurements and discuss the observed effects and their causes.

3:05—3:20 Break

3:20

1pSP8. Acoustic and direct vibratory excitation, propagation, and scattering of Rayleigh waves at an air-ground interface. Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, One Coliseum Dr., University, MS 38677), R. Daniel Costley (Miltec Res. and Technol., Oxford, MS 38655), Craig J. Hickey, and James M. Sabatier (Natl. Ctr. for Physical Acoust., University, MS 38677)

Finite element methods are utilized to model loudspeaker as well as shaker generation of sound, shear, and interface waves in an elastic solid containing an imbedded elastic scatterer. Results for steady state (continuous) and transient (pulsed) insonification, animated in the time domain, are presented to illustrate excitation, propagation, and scattering mechanisms and features. [Work supported by the U.S. Army Research Development and Engineering Command.]

3:35

1pSP9. Estimation of pointing vector using acoustic planar array. Thyagaraju Damarla (RDRL-SES-A, US Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, thyagaraju.damarla@us.army.mil)

Traditional use of acoustic array of sensors for estimation of azimuth and elevation of a target suffer from inaccuracies, especially in estimation of elevation angle, due to poor resolution in trigonometric functions. For example, a variation of angle 70×90 deg result in variation of sine of angle in 0.94–1. In the proposed approach, each pair of elements (sensors) in the array is used to estimate the direction of arrival (DoA) of the signal. This DoA forms a cone along the axis of the line joining the pair of sensors. Several such cones from different pairs of sensors would intersect giving the direction of the target. Since the phase angles between two sensors can be estimated accurately, the DoAs will be accurate and hence the intersection of the cones would result in accurate estimation of the pointing vector to the target overcoming the deficiencies of the traditional methods. In the case of transient events, the time difference of arrival at the pair of sensors gives the DoA. Mathematical formulation for estimating the pointing vector will be given. The technique will be used on some field data and the results obtained will be compared with the ground truth.

3:50

1pSP10. Near-ultrasound, broadband noise generated by low-speed airflow over wires. W.C. Kirkpatrick Alberts, II, Kevin J. Sanchez, Mark A. Coleman, Josh Gabrielse, and David A. Ligon (US Army Res. Lab., ATTN: RDRL-CIE-S, 2800 Powder Mill Rd., Adelphi, MD 20783)

Vortex shedding due to flow past a cylindrical object and its accompanying sound are well-described phenomena, though research has often focused on the sound broadcast in directions transverse to the wire and downstream in high-speed flows. Here, we investigate the sound broadcast in angles nearly opposite the direction of low-speed flows from nozzles of diameter similar to those of various wires. Also investigated are the effects that several parameters (tension, wire diameter, nozzle-wire separation, etc.) have on the level and frequency content of the recorded sound. It is found that a microphone approximately 3 cm above the outlet of a nozzle records broadband, near-ultrasound noise due to the flow over the wire. Signal levels as high as 20 dB above the nozzle background are observed. Recorded broadband noise spectra associated with vortex shedding from various wires and schlieren images of the interaction between the flow and the wire will be presented.

4:05

1pSP11. Matched field processing applied to diffraction of sound around 90 deg corners: Modeling and experiment. Victor Singh (W. E. Lay Auto Lab., Univ. of Mich, 1697 Broadway St. 303, Ann Arbor, MI 48105, victorsi@umich.edu), Katherine E. Knisely, Karl Grosh, David R. Dowling (Univ. of Mich, Ann Arbor, MI 48105), and Serdar H. Yonak (Toyota Motor Engr. and Mfg. North America, Ann Arbor, MI 48105)

Acoustic diffraction allows sound to travel around opaque objects and therefore may allow beyond-line-of-sight sensing of remote sound sources. This presentation reports simulated and experimental results for detecting and localizing sound sources at blind city-street intersections based on shadowed microphone array measurements. The generic geometry includes a point source, a solid 90 deg wedge, and a receiving array that lies entirely in the shadow defined by the source location and the wedge. Sound source detection and localization performance are assessed via matched-field (MF) ambiguity surfaces as a function of source frequency, receiving-array configuration, and received signal-to-noise ratio for the Bartlett and minimum variance distortionless MF processors. Here, the sound propagation model is developed from a Green's function integral treatment and the geometric theory of diffraction. The simulations suggest that sound sources may be localized by fully shadowed arrays for signal-to-noise ratios as low as 20 dB when ambiguity surfaces from several different source frequencies are incoherently combined. The experiments were conducted using an approximate 50-to-1-scaled tabletop model of a blind city-street intersection and chirp signals from 500 Hz to 30 kHz and suggest that the propagation model must be improved before the simulations' performance can be reached.

4:20

1pSP12. Revisiting isodiachrons as a solution for localization. Balaji Katlai and Jonathan Vallarta (Jasco Appl. Sci., 1496 Lower Water St., Halifax, NS, Canada)

Recently, there have been some new attemtps in source localization algorithms. In particular, Spiesberger *et al.* successfully introduced the notion of isodiachrons to localize a source. In this abstract, it is shown that the concept and notion of using an isodiachron can be understood in terms of mapping an unbounded trajectory (as in conventional hyperbolic fixing for localization) into a more closed form bounded surface. This is made possible by realizing a simple analytical continuation of the parameters in the algorithm. The resulting system is a bounded surface with isodiachrons as a solution. A simple proof-of-concept example is also demonstrated. This is a complementary solution/approach to the original idea of isodiachrons as proposed by Spiesberger. Additionally, other alternative approaches to translate (an unbounded) hyperbolic fixing technique to a source on a bounded surface are also discussed.

4:35

1pSP13. Optimal selection of weights for the linear total least-squares solution to impulsive source localization. Wm. Garth Frazier and Heath Rice (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University MS 38677)

At a recent ASA meeting, Monte Carlo-based noise performance analysis of several formulations and solution algorithms for the time-of-arrival (TOA) based impulsive source localization problem were presented. That analysis included two algorithms for parameter estimation using the (exact) linear formulation for the localization problem. Those algorithms were standard linear least-squares (LLS) and linear total least-squares (TLS). Using systematically derived, but in no way optimized, choice of weights, the TLS algorithm clearly demonstrated the ability to overcome the bias associated with LLS under conditions associated with low signal-to-noise ratio. Unfortunately, the bias reduction was obtained at a cost of high variance as compared to solutions associated with nonlinear formulations. This presentation presents results from investigations into performance of optimally chosen (variance reducing) weighting schemes for TLS and compares them to the performance of nonlinear formulations.

4:50

1pSP14. Error analysis of locating arbitrary sound sources in three dimensional space in real time. Na Zhu and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202, dv6930@wayne.edu)

This paper presents an error analysis in locating multiple incoherent sound sources in three dimensional space in real time by using a newly developed hybrid approach. The underlying principle of this hybrid approach consists of modeling of sound radiation from point sources in a free field, triangulation, and signal processing techniques. The impacts of target source locations and signal to noise ratio on the results are examined. Experimental results show that errors in source localization consist of primarily two components: biased and random errors. The former seems to be caused by the fact that the formulations used in acoustic modeling are for a free field, whereas the measurement environment is a non-free field. The latter is mainly caused by fluctuations in the source levels. To account for these errors, empirical formulations are developed and validated again in experiments with different settings. Results demonstrate that these empirical formulations can significantly improve the accuracy and spatial resolution of source localizations.

5:05

1pSP15. Coherence of low-frequency acoustic signals across a tetrahedral array. Geoffrey, H Goldman (U.S. Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783-1197)

Many direction finding algorithms for acoustic applications require that the data collected at each microphone is coherent with respect to the data collected with another microphone in some frame. As the microphones are moved farther apart, the angular resolving power of the array is increased, but the coherence of the data is reduced. These factors are critical for designing the spatial configuration of acoustic arrays. The coherence of low-frequency acoustic signals was examined across tetrahedral microphone arrays elevated 1 m above the ground. The signal source was a UH-1 helicopter, which had strong harmonics between 10 and 100 Hz. The coherence was measured by tracking the harmonics of the signal, bandpass filtering each harmonic, cross correlating the filtered signals across all the microphone combinations, finding the maximum output, and then normalizing. The results indicate that the elevated microphone was less coherent than the non-elevated microphones. Multipath was determined to be a major factor affecting the coherency.

Session 1pUW

Underwater Acoustics and Acoustical Oceanography: Deep Water Ambient Noise

Michael J. Buckingham, Cochair
Univ. of California, San Diego, Scripps Inst. of Oceanography, 9500 Gilman Dr., La Jolla, CA 92093-0238

Martin Siderius, Cochair
Portland State Univ., Dept. of Electrical and Computer Engineering, 1900 SW 4th Ave., Ste. 160-11, Portland, OR 97201

Chair's Introduction—1:05

Invited Papers

1:10

1pUW1. Oceanic noise: Mechanisms, radiation characteristics, and array results. William M. Carey (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Noise produced by the oscillation of microbubble distributions and impacts (raindrops and splash) in the air-sea boundary zone results in doublet radiation patterns and noise fields with definitive temporal and spatial characteristics. Turbulence generated noise is quadrupole, and nonlinear wave-wave interactions are infrasonic. Plausible mechanisms provide a framework for a review of directional noise measurements in range dependent oceanic and bathymetric environments. Slope interaction converts higher angles to lower angles and results in a frequency dependent vertical directionality. Low-frequency (0.02–1 kHz) measurements reveal a temporally dynamic noise field with persistent directional characteristics. Higher frequency (1-kHz or greater) measurements exhibit a local stationary characteristic influenced by the boundary zone mixed layer. Beam noise cumulative distribution functions are shown to depend on the distribution of ships, basin size, and boundaries. Directional noise computations are presented along with the issues of shipping distributions, the air-sea boundary zone, and assimilated satellite observations of wind speed, white caps, and sea surface temperature. The computation of basin scale directionality basically estimates the persistent characteristic. Improved calculation with the directional radiation from modern ships and the inclusion of the dynamic ship distribution ships are presented to estimate the beam noise statistics.

1:30

1pUW2. Trends in low-frequency deep ocean ambient noise levels: New results from old data. Ross Chapman and Andrea Price (School of Earth and Ocean Sci., Univ. of Victoria, 3800 Finnerty Rd., Victoria, BC V8P5C2, Canada, chapman@uvic.ca)

Concern about effects of anthropogenic sound generated by ships, seismic exploration, naval operations, and ocean acoustic research on marine life has stimulated new studies of marine animal bioacoustics. An underlying issue in assessing the noise impact establishes the present-day background noise level and compares it to noise levels from previous times. The most generally accepted prediction of noise level trend at low frequencies (below 500 Hz) is due to Ross, who used noise levels from the 1950s and the mid-1960s to predict an increase of about 3 dB/decade. This paper addresses the issue of the trend in low-frequency ambient noise levels in the ocean and presents measurements taken in the intermediate years between the mid-1960s and the present time. The measurements were made using a calibrated multi-element volume array (MEVA) at deep ocean sites in the Northeast Pacific from 1978 to 1986. The array provided spectral noise levels and horizontal and vertical directionalities. Compared to the 1965 values, the MEVA values are 6–8 dB greater between 10–50 Hz and 1–3 dB greater from 50–400 Hz, respectively. The data presented here provide evidence that the trend predicted by Ross extended at least into the mid-1980s.

1:50

1pUW3. Noise cross-correlation and passive measurements of the sound speed profile in deep water. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., Boulder, CO 80305-3328, oleg.godin@noaa.gov), Nikolay A. Zabotin (Univ. of Colorado, Boulder, CO 80309-0425), and Valery V. Goncharov (Russian Acad. of Sci., Moscow 117997, Russia)

Ambient acoustic noise in the ocean contains extensive information about its sources and the propagation environment. Successful application of the noise observations to acoustic characterization of the environment depends on identifying the noise properties which can be reliably measured and are sensitive to variations in temperature and other physical parameters of the ocean. This paper focuses on two-point noise cross-correlation and the environmental information that can be retrieved from it without any *a priori* knowledge about properties and locations of the noise sources. A technique to retrieve deterministic acoustic travel times from cross-correlations of noise recorded on two vertical line arrays is described. Feasibility of environmental monitoring with ambient noise is illustrated by results of passive tomography of the water column using noise recordings of opportunity, which were obtained in the North Pacific Ocean as a by-product of a long-range sound propagation experiment. Conflicting requirements on the hydrophone separation and frequency band are analyzed that originate from the demands to improve the inversion accuracy and to decrease noise averaging time. Remaining challenges in retrieving oceanographically meaningful results from acoustic noise cross-correlations are discussed. [Work supported by ONR.]

2:10

1pUW4. Depth-profiling ambient noise in the deep ocean. David R. Barclay, Fernando Simonet, and Michael J. Buckingham (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238, mbuckingham@.ucsd.edu)

Deep Sound is an untethered instrument platform designed to free-fall from the sea surface to a preassigned depth, at which point a burn wire releases a weight, allowing the system to return to the surface under buoyancy. The descent and ascent rate is 0.6 m/s. A Vitrovex glass sphere houses lithium-ion batteries and a suite of microprocessor-controlled electronics for data acquisition, data storage, power management, and system control. Outside the sphere, several hydrophones are arranged in vertical and horizontal configurations, a CTD returns environmental data, and motion sensors monitor pitch, roll, and yaw. Data may be downloaded, and the batteries may be recharged, via throughputs in the sphere. The hydrophones, with a bandwidth of 30 kHz, are rated to a depth exceeding 11 km, and the sphere itself has a depth-rating of 9 km. The system made three descents in the Philippine Sea in May 2009, to depths of 5100, 5500 and 6000 m; and in November 2009, two descents were made in the Mariana Trench to a depth of 9000 m. On all these deployments, ambient-noise time-series were recorded continuously, yielding the power spectrum and vertical coherence of the noise as functions of depth. [Research supported by ONR.]

Contributed Papers

2:30

1pUW5. The statistical interpretation of a simple ambient noise model. Richard Evans (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, richard.evans.01@snet.net)

Theoretical analyses of the statistics of ambient noise levels can yield distributions that are either narrow and symmetric, or broad and nonsymmetric [I. Dyer, J. Acoust. Soc. Am. **53**, 564–570 (1973)]. Their standard deviations vary between nearly 0 and 5.6 dB. The assumptions regarding the nature of the noise source are the key in determining what distribution to expect. A simple computational model that exhibits both extremes, of the potential statistical distributions, can help in understanding what assumptions determine the statistics. The paper presents a simple ambient noise model that allows such a statistical interpretation.

2:45

1pUW6. Statistical properties of deep ocean noise. Brianne Moskovitz, Gerald D'Spain, John Hildebrand, and Achintya Madduri (Marine Physical Lab., Scripps Inst. of Oceanogr., 291 Rosecrans St., San Diego, CA 92106, bmoskovi@ucsd.edu)

The various arrays sea test (VAST) was conducted in July 1989 in the northeast Pacific Ocean. A 3000-m-long, 200-element hydrophone array was suspended for an 11-day period from R/P FLIP, moored about halfway between San Diego and Hawaii (34N, 140W). Simultaneously, 12 neutrally buoyant, freely drifting Swallow floats equipped with both infrasonic hydrophones and three-component geophones were deployed at various depths 150 km off the California coast (35N, 122W). The statistical properties of deep ocean noise recorded by these sensor systems, both single element and beam level, are examined quantitatively under a variety of conditions, including wind-dominated (15–18 kt), distant air gun operations, and marine mammal vocalizations. Non-parametric statistical tests applied to narrow band complex time series of the data include the Wald–Wolfowitz run test for mutual independence of the data samples, the Kolmogorov–Smirnov two-sample test for stationarity, and the Kolmogorov–Smirnov one-sample test for Gaussianity. Results are presented over a broad range in frequencies and time scales and show that the noise field at times and in certain directions displays non-stationary and non-Gaussian properties. In a few cases, these statistical characteristics can be associated with specific physical processes. [Work supported by the Office of Naval Research.]

3:00—3:15 Break

Invited Papers

3:15

1pUW7. Deep water ambient noise on the seafloor. Ralph A. Stephen (Dept. of Geology and Geophys., Woods Hole Oceanograph. Inst., 360 Woods Hole Rd., Woods Hole, MA 02543, rstephen@whoi.edu)

Long, continuous time series of ocean bottom seismometer (OBS) and co-located hydrophone data from the seafloor offer an unprecedented opportunity to study the physical mechanisms of seafloor ambient noise in the frequency band from 1 to 100 Hz. The Ocean Seismic Network Pilot Experiment (OSNPE) in 1998 compared noise and signal levels for a hydrophone at the seafloor with three-component observations from OBSs on the seafloor, buried 1 m into the soft sediment, and clamped in a borehole 10 m into basaltic basement. The Hawaii-2 Observatory (H2O) acquired over 3 years of nearly continuous, real-time data from a shallow-buried seismic and seafloor hydrophone system. On the North Pacific Acoustic Laboratory experiment in 2004 (NPAL04) three vertical-component geophone/hydrophones were deployed beneath vertical line arrays in the water column. In addition to ambient noise data, controlled source transmissions were received from ranges out to 3200 km. In the absence of shipping noise, whale calls, and local sea state, what controls the noise floor on the deep seafloor? Interesting observations include banded coherence structure between the vertical geophones and hydrophones, peaks in ambient noise levels due to resonances in the sediment layer, and Stoneley/Scholte wave particle motions. [Work supported by the NSF, ONR and WHOI.]

3:35

1pUW8. Long-time trends in low-frequency traffic noise for four sites off the North American west coast. Rex K. Andrew (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, randrew@apl.washington.edu), Bruce M. Howe (Univ. of Hawaii at Manoa, Honolulu, HI 96822), and James A. Mercer (Univ. of Washington, Seattle, WA 98105)

Measurements from four cabled-to-shore hydrophone systems located off the North American west coast permit extensive comparisons between "contemporary" low-frequency traffic noise (25–50 Hz) collected in the past decade to measurements made in the mid-1960s with the same in-water equipment at the same sites. An increase of roughly 10 dB over the band 25–40 Hz at one site has already been reported [Andrew *et al.*, ARLO Acoust. Res. Let. Online **3** (2002)]. Newly corrected data from the remaining three systems corroborate this increase. Simple linear trend lines of the contemporary traffic noise (duration 6–12+ years) show that the current levels either hold steady or decrease at three of the four sites. These results confirm the Ross prediction, at least at these sites, that the rate of increase in traffic noise would be far less at the end of the last century compared to that observed in the 1950s and 1960s.

3:55

1pUW9. Deep ocean noise measurements in the Philippine Sea. Matthew Dzieciuch and Peter Worcester (SIO/UCSD, 9500 Gilman Dr., La Jolla, CA 92039, mad@ucsd.edu)

In the spring of 2009, a newly developed distributed vertical line array (DVLA) receiver made up of two, 1000-m-long subarrays with 30 hydrophones each was deployed for approximately 1 month in the northern Philippine Sea. One subarray spanned the sound-channel axis, and the second subarray spanned the surface conjugate depth (about 4200 m). Twenty of the 30 hydrophones in the deep subarray were spaced 5 m apart (half-lambda at 150 Hz), allowing the vertical directionality of the ambient noise below the surface conjugate depth to be determined. The minimum omnidirectional ambient noise levels decreased significantly below the surface conjugate depth at frequencies from 50 to 500 Hz, similar to behavior previously observed in the central North Pacific Ocean. The minimum noise levels presumably correspond to times when wind speeds are low and surface conditions are calm so that there is little locally generated, low-frequency noise.

Contributed Papers

4:15

1pUW10. Model-data comparisons of deep water ambient noise. James J. Murray (OASIS, Inc., Fairfax, VA), Peter Worcester, Matthew Dzieciuch (Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA), and Kevin D. Heaney (OASIS, Inc., Fairfax, VA)

During the North Pacific Laboratory (NPAL) Philippine Sea 2009 experiment, a distributed vertical line array (DVLA) was deployed for several weeks in the northern Philippine Sea. The array, described in detail in a previous talk, consisted of two 1000-m sections, one spanning the sound channel axis and the other covering the deep ocean, extending from above the surface conjugate depth down to within roughly 100 m of the bottom. In this paper, a hybrid propagation model for handling low-frequency ocean noise will be presented. The model uses analytic solutions (rays) for ranges to 5 km, the parabolic equation for ranges from 5 to 100 km, and normal modes for ranges beyond that. The model predicts ambient noise levels vs depth. The fall off of the ambient noise with depth below the surface conjugate has been reported. The level of the fall off and its frequency dependence are sensitive to the local geo-acoustics. A geo-acoustic inversion will be performed using a range-independent version of the acoustic model with Monte-Carlo realizations of surface ship distributions.

4:30

1pUW11. Horizontal noise directionality in deep water measurements and modeling. Richard Campbell, Kevin D. Heaney (OASIS, Inc., Fairfax, VA 22039), and Arthur B. Baggeroer (MIT, Cambridge, MA)

Low-frequency, less than 100 Hz, sound is dominated by surface shipping, particularly in the northern hemisphere. With low-volume attenuation sound can propagate significant distances. For periods of time not dominated by a particular nearby interferer, the low-frequency ambient noise field is dominated by distant shipping and is expected to be anisotropic due to various bathymetric effects such as downslope conversion and seamount or ocean ridge blockage. In this paper, the authors present horizontal array measurements from a 100 m towed line array (towed at 100 m) in very deep water in the Philippine Sea. The northern Philippine Sea has several ridges and seamounts and significant bathymetric interaction for surface ships in the main Singapore Shanghai shipping lane. A rapid range bearing dependent $N\times2$D parabolic equation approach will be described and used to evaluate the bathymetric effects on the horizontal directionality of low-frequency noise. Comparisons of measurements and models will be done using both Monte-Carlo realizations of surface ship positions (from the HITS database) as well as measured AIS from research vessels in the vicinity.

4:45

1pUW12. Characteristics and variability of ambient noise in the South China Sea basin. Ching-Sang Chiu, John E. Joseph, and Christopher W. Miller (Dept. of Oceanogr., Naval Postgrad. School, 833 Dyer Rd., Rm. 328, Monterey, CA 93943)

A hydrophone was moored at mid depth in the South China Sea basin from November 2005 to October 2006. Operated with a 1-min-on and 14-min-off duty cycle and sampled at 1.6 kHz, the measured time series captures the spectral characteristics and variability of the ambient noise in the 0–800-Hz band over an annual cycle. In this paper, we provide a description on the daily, monthly, and seasonal variabilities and variances in the measured noise spectrum and band levels. In order to gain insights into the predictability of the ambient noise field in this marginal sea, the interpretation of the data is facilitated with historical shipping density data and nowcast wind fields from the Navy Operational Global Atmospheric Prediction System (NOGAPS). Attention is made to the evolution of the ambient noise spectra during major storm events. Intermittent noises are also examined. The potential sources for these intermittent noises are discussed. [Research sponsored by the Office of Naval Research.]

5:00

1pUW13. Blue whale vocalizations and the seasonal variability of ambient noise levels at the Aloha Observatory. Orest Diachok (Appl. Phys. Lab., Johns Hopkins Univ., Laurel, MD) and Fred Duennebier (Univ. of Hawaii, Honolulu, HI)

In February, 2007 the Aloha Observatory added a cabled, bottom mounted hydrophone, which supplied a 20-month continuous recording of ambient noise from 0.02 Hz to 11 kHz. This observatory is situated at a depth of 4700 m, 100 km north of Oahu. This site is known to be close to (possibly within) one of the primary habitats of NE Pacific and NW Pacific blues, and other whales during the winter. Numerous previously reported measurements throughout the NE Pacific Ocean have demonstrated that the fundamental frequencies of NE blues vocalizations have systematically decreased from about 22.5 Hz in 1958 to about 16.5 Hz in 2007 (the cause of the frequency shift is unknown). Limited measurements suggest that the fundamental frequencies of NW blues vocalizations have also decreased with time and were about 18 Hz in 2007. Ambient noise levels from the Aloha site were analyzed in 1-Hz bands for 1 year to permit observation of the seasonal dependence of ambient noise levels as a function of frequency. The data reveal distinct peaks in noise levels at 16.5 and 18.5 Hz between November and March, consistent with observations of the fundamental frequencies of NE and NW blue whale vocalizations. Since blues and other whales repertoire of vocalizations cover a broad frequency range, the contributions of their vocalizations to noise levels at other frequencies will be considered.

5:15

1pUW14. Invariant features of deep ocean ambient noise. Zachary S. Guralnik (SAIC, 4001 N. Fairfax Dr., Ste. 175, Arlington, VA 22203)

Very low-frequency, deep ocean ambient noise is generated by nonlinear interactions (collisions) between surface waves with nearly equal frequency, propagating in nearly opposite direction. It is shown that the usual "standing wave" approximation used to compute this noise predicts that the ratios of entries in the power spectral density matrix, obtained from auto and cross-correlations of acoustic velocity and pressure fluctuations, are universal constants. This prediction is too strong and not observed. A careful consideration of bottom effects invalidates the standing wave approximation. Nevertheless a weak version of the standing wave approximation still holds and has interesting implications. While the ratios of entries in the power spectral density matrix due to surface wave generated noise are not universal constants, they are insensitive to the details of the surface wave spectrum. This theoretical prediction is borne out by data from the Hawaii-2 Observatory. [Work supported by ONR.]

Separate registration fee required

MONDAY AFTERNOON, 19 APRIL 2010 HARBORSIDE B, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Animal Hearing

Micheal L. Dent, Chair
SUNY Buffalo, Dept. of Psychology, Buffalo, NY 14260

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Animal hearing. Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

The auditory world of animals is, in many cases, quite different from our own due to evolutionary pressures that have created anatomical and physiological differences in auditory structures. However, there are a number of commonalities of function and mechanisms across many species that point to general concepts of auditory perception. This tutorial lecture explores the field of animal hearing, also known as comparative psychoacoustics. Because the diversity in hearing organs among animals is considerable, the hearing of many animals is quite different from our own. We will review some of these differences in hearing across animal groups, from household pets to more exotic animals. This tutorial also highlights the specific advantages of the comparative approach and illustrates the many creative methods used for behavioral testing of hearing in animals. Species comparisons contribute to our understanding of the evolution of the auditory system and can often clarify the relationship between structure and function in the auditory system as well as the effects of damage and repair. The study of animal hearing can lead to improved understanding of human auditory capabilities and invites speculation about how human speech is adapted for intraspecies communication.

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Environmental Effects on Acoustic Propagation

Roger M. Oba, Cochair
Naval Research Lab., Code 7120, Washington, DC 20375

Ying-Tsong Lin, Cochair
Woods Hole Oceanographic Inst., 210 Bigelow Bldg., Woods Hole, MA 02543

Contributed Papers

9:15

2aAO1. Acoustic ducting and refraction by sea bottom relief in shallow water. James F. Lynch, Alexey A. Shmelev, Ying-Tsong Lin, and Arthur E. Newhall (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 02543, jlynch@whoi.edu)

Observations show that shallow water bottom relief often has a band-limited directional spectrum produced by various oceanographic and geological processes. This directional bottom feature is shown to have a noticeable effect on three-dimensional low-frequency acoustic propagation. An analytical study with an idealized model of straight sea bottom ripples has shown that acoustic energy can be partially ducted between neighboring ripples, and this ducting will affect acoustic propagation in shallow water. In our work, we also study ducting and refracting due to idealized curved sea bottom ripples. Previous research has shown that non-linear internal waves can also create acoustical ducts. Comparative analysis of these two different ducts is performed using our idealized model. The combined effects of internal waves and bathymetry are studied for various relative directions of internal wave front and bottom ripples. A numerical simulation of three-dimensional sound propagation across realistic bathymetry and internal wave fluctuations is performed. In conclusion, both water column fluctuations and bathymetry variability need to be taken into account when studying three-dimensional acoustic propagation in shallow water.

9:30

2aAO2. Three-dimensional sound propagation over submarine canyons. Ying-Tsong Lin, Timothy F. Duda (Dept. Appl. Ocean Phys. & Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, ytlin@whoi.edu), Jon M. Collis (Colorado School of Mines, Golden, CO 80401), James F. Lynch, and Arthur E. Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Submarine canyons are common features of continental shelf and slope regions, e.g., Hudson Canyon in the Mid-Atlantic Bight. In this paper, the impact of submarine canyons on low-frequency sound propagation is studied using a three dimensional (3-D) parabolic approximation numerical program, which is implemented in a Cartesian coordinate system and utilizes the split-step Fourier technique and a 3-D variant of the Thomson and Chapman wide-angle approximation. This program will be first benchmarked with a classic wedge problem, and then used to study an idealized canyon environment to understand distinct 3-D sound propagation effects. The idealized environment has a Gaussian shaped canyon incising a slope. Horizontal focusing of sound in the canyon and energy flow into the canyon from an off-axis sound source are observed. A realistic model using the Hudson Canyon bathymetry shows even more complex sound propagation situations. Propagation conditions over different seabed types are also compared, and the 3-D field sensitivity to bottom properties is investigated. [Work supported by the Office of Naval Research.]

9:45

2aAO3. Three dimensional parabolic equation modeling of an internal wave event during Shallow Water 2006. Georges A. Dossot, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett Bay Campus, Narragansett, RI 02882), James F. Lynch, Ying-Tsong Lin (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), and Kevin, B. Smith (Naval Postgrad. School, Monterey, CA 93943)

During the Shallow Water 2006 (SW06) experiment, a J-15 acoustic source deployed from the Research Vessel Sharp transmitted broadband (100–500 Hz) chirp signals 15 km away from a vertical line array. The array was intentionally positioned near the shelf-break front and in an area where internal waves are known to occur. During the same time an internal wave, "Event 44," passed through the sound field such that the internal wave front was near parallel to the acoustic transmission path. Measured data show substantial intensity fluctuations that vary over time and space due to complex multimode and multipath (both two and three dimensional) interference patterns. Of specific interest are fluctuations of measured intensity preceding the internal wave's arrival. Additionally, depth variability of the measured acoustic intensities can be attributed to a warm water intrusion coinciding with the internal wave event. This presentation shows recent modeling results using the experimental geometry, acoustic signal parameters, and a simulated oceanographic environment based on environmental moorings and ship-born sensors. A new version of the three-dimensional Monterey–Miami parabolic equation code, which incorporates a user-defined sound speed field, is used. [Work sponsored by the Office of Naval Research.]

10:00

2aAO4. Horizontal focusing/defocusing due to shallow-water internal waves. Jing Luo, Mohsen Badiey (Univ. of Delaware, Robinson Hall 112B, 261 S. College Ave., Newark, DE 19716, luojing@udel.edu), and Ying-Tsong Lin (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

During the New Jersey Shallow Water 2006 (SW06) experiment, an acoustic source was towed by the Research Vessel Sharp and followed the front of an internal wave packet. The source was transmitting broadband acoustic signals (50–450 Hz) in different angles with respect to the internal wave front. The receptions of transmitted signal on a vertical hydrophone line array are analyzed to study the horizontal focusing/defocusing that occurred when the internal wave front and acoustic track aligned closely. Based on ship-board radar images and temperature data collected on the environmental moorings at various locations along the acoustic track, a detailed three-dimensional (3D) environment is reconstructed for a 3-D parabolic approximation model to study the unique propagation scenario. Construction of index of normal mode refraction for these data provides a clear picture of acoustic energy focusing for this event. Data and model comparison are in good agreement. [Work supported by ONR321 OA]

10:15—10:30 Break

10:30

2aAO5. Spatial and temporal sound fluctuations in shallow water in presence of internal soliton in transition areas. Boris Katsnelson and Andrey Malikhin (Voronezh Univ., 1 Universitetskaya sq, Voronezh 394006, Russia)

Behavior of the sound field is considered in a 1–2-h time interval in presence of train of internal soliton (IS) crossing an acoustic track. During this period moving ISs pass through several stages: approaching an acoustic track, consequent covering source (receiver) only, both source and receiver, receiver (source) only, and receding from an acoustic track. Correspondingly there are different regimes of interaction of the sound field with IS: horizontal reflection, capture of signals in horizontal waveguide, adiabatic variations, and transition areas between these regimes. So, rather complex spatial and temporal structure of the sound field takes place, including frequency and modal dependence of its parameters. Theoretical analysis of variations in the sound field on the basis of techniques of vertical modes and horizontal rays (PE in horizontal plane) is carried out, and estimation of feasibility of an experimental setup is presented. [Work was supported by CRDF.]

10:45

2aAO6. Calculating the waveguide invariant for non-ideal waveguides. Kevin L. Cockrell and Henrik Schmidt (Dept. of Mech. Eng., MIT, Cambridge, MA 02139)

The waveguide invariant describes striations in a range versus frequency plot of a waveguide's Green's function. Analytic expressions for the waveguide invariant only exist for a few select waveguides, but experiments and simulations have shown that the waveguide invariant is approximately equal to unity for almost all realistic shallow-water waveguides. A quasi-analytic method will be presented for estimating the value of the waveguide invariant in waveguides with arbitrary sound speed profiles, including the effects of a bottom fluid halfspace. The method is approximate but allows for an intuitive understanding of why the value of the waveguide invariant does not strongly depend on the details of the sound speed profile.

11:00

2aAO7. Arrival structure variability of single-bounce paths for high-frequency transmissions during the experiment KAM08 (Kauai acoustics communications multidisciplinary research initative 2008). Joseph M. Senne, Aijun Song (Univ. of Delaware, 210 Robinson Hall, Newark, DE 19711, sennejm@udel.edu), and Kevin B. Smith (Graduate School of Eng. and Appl. Sci., Monterey, CA 93943)

During the summer of 2008 an experiment was conducted that included both chirp and M-sequence transmissions at 16-kHz center frequency. Source and receiver arrays were located west of Kauai Island HI, along an isobath of about 100 m. Moored thermistor strings and a wave-rider buoy provided detailed oceanographic data, while shipboard measurements recorded wind variations. Micro-multi-paths were observed from single surface bounces, and their variability has been examined for a variety of surface wave conditions. A surface wave model has been integrated into a parabolic equation model (MMPE) to approximate variations in the micro-multi-path structure over geotime. Model results are used to examine the correlation between environmental variability and observed single-bounce signal fluctuations. Comparisons are made with a variety of surface wave conditions, including both calm and rough seas. [Work supported by ONR 321OA.]

11:15

2aAO8. Impact of surface gravity waves on high-frequency acoustic propagation in shallow water. Entin A. Karjadi, Mohsen Badiey (College of Earth, Ocean, and Environment, Univ. of Delaware, 210 Robinson Hall, Newark, DE 19716, karjadi@udel.edu), and James T. Kirby, Jr. (Univ. of Delaware, Newark, DE 19716)

Sea surface roughness is one of several factors that significantly influences high-frequency (1–50-kHz) acoustic wave propagation in shallow water. The evolving sea surface introduces several variability effects including Doppler shift. Data analyses from high-frequency acoustic experiments show high-correlation between time, angle, and intensity fluctuations of received signals and varying sea surface conditions. In order to assess detailed acoustic signal interactions with the sea surface, a realistic wave model is developed and combined with an acoustic ray-based model. Model validity is evaluated by comparing the results with data from multiple experiments. [Work supported by ONR 321OA.]

11:30

2aAO9. Acoustic observations of subsurface instability. Justin M. Eickmeier, Mohsen Badiey (College of Earth, Ocean and Environment, Univ. of Delaware, Newark, DE 19716, jeickmei@udel.edu), and Tokuo Yamamoto (Rosenstiel School of Marine and Atmospheric Sci., Univ. of Miami, Miami, FL 33149)

A high-frequency (0.6–18-kHz), shallow water acoustic experiment (HFA2000) was conducted in Delaware Bay (15-m depth) during December 2000. Reciprocal transmissions of chirp signals (0.345-s duration) were radiated between three bottom mounted source-receiver tripod stations separated by 70–353 m. Environmental data were collected at a nearby oceanographic platform; simultaneously, a shipboard ADCP and CTD were deployed. Analysis of direct path station-to-station arrival times (between December 18th 00:00 and December 19th 10:00, during which 126 acoustic transmissions consisting of 29 chirps were radiated) revealed significant deviation from arrival time patterns established during previous tidal cycles. Examination of the corresponding signal intensity reflected this deviation. Independent ADCP data displayed current profile distortion during the period along the direction of the dominant flow channel. The mean slope of a wave number vs geo-time spectrum was calculated from each geo-time's respective chirp series. Slope changes correlate to variations in the total signal intensity's constituents, $I_{\mathrm{tot}}(t) = \langle |E(t)|^2 \rangle$, particularly the incoherent or scattered intensity, $I_{\mathrm{incoh}}(t) - I_{\mathrm{total}}(t) - I_{\mathrm{coh}}(t)$. Through comparison with a Kolmogorov power spectrum and calculation of the corresponding Richardson number, a profile of environmentally induced subsurface instability has been developed.

11:45

2aAO10. The three-dimensional acoustic field of primary arrivals from a seismic airgun array. Arslan M. Tashmukhambetov, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, atashmuk@uno.edu), Natalia A. Sidorovskaia, Anca Niculescu (Univ. of Louisiana at Lafayette, Lafayette, LA), Joal J. Newcomb (Naval Oceanograph. Office, Stennis Space Ctr., MS), James M. Stephens, Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS), and Phil Summerfield (ExxonMobil Corp., Houston, TX)

The Source Characterization Study 2007 (SCS07) measured the three-dimensional (3-D) acoustic field of a seismic airgun array. The Littoral Acoustic Demonstration Center (LADC) performed the experiment, collecting acoustic and related data on three moored hydrophone arrays and one ship-deployed hydrophone array which together spanned the full water column. Sensitive and desensitized phones were deployed at each position to extend the dynamic range. An ultra short baseline localization system was deployed with the EARS moorings to provide array shape. With postanalysis this results in time-dependent positions for each of the acoustic sensors. Every channel is calibrated. A seismic source vessel shot a series of lines designed to give detailed angle and range information concerning the field of the primary arrival. Peak pressures, sound exposure levels, total shot energy spectra, and one-third octave band analyses give important insights into details of the acoustic field. Images of these quantities are generated to show dependence on emission and azimuthal angles and range. 3-D visualizations and two-dimensional cuts through the data are shown. [Research supported by the Joint Industry Programme through the International Association of Oil and Gas Producers.]

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration, Animal Bioacoustics, Speech Communication, and Psychological and Physiological Acoustics: Blast-Induced Traumatic Brain Injury: Mechanisms, Assessment, Therapy, and Mitigation

Steven G. Kargl, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

William C. Moss, Cochair
Lawrence Livermore National Lab., 7000 East Ave., Livermore, CA 94550

Thomas J. Matula, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Chair's Introduction—8:05

Invited Papers

8:10

2aBB1. Defense Advanced Research Projects Agency (DARPA) Blast Program–Preventing violent neurologic trauma. Geoffry Ling (Defense Advanced Res. Projects Agency, 3701 N. Fairfax Ave., Arlington, VA 22203)

DARPA is directing a comprehensive program to determine the causes of explosive neurological trauma. The current status of the program will be discussed.

8:30

2aBB2. Basics of blast physics. David Ritzel (19 Laird Ave. North, Amherstburg, ON N9V 2T5, Canada)

Blast science concerns the processes by which the energy of an explosion source becomes propagated into its surrounding environment, then interacts, loads, and damages materials, structures, and systems. Although explosive events have been chronicled for over 2000 yrs, and good empirical insights regarding blast phenomenology were established prior to WW-II, the advent of the nuclear bomb drove the first rigorous and concerted efforts to understand the detailed physics of blast propagation. Blast science merges with classical acoustic sciences in its far-field limits including phenomena such as atmospheric focusing. However, in the interaction of blast with composite structures including the human body, a wide range of transmitted, coupled, or overdriven mechanical waves of many types can be generated. There has been a significant resurgence of R&D in protective technologies against modern blast threats such as terrorist bombings and IED attacks against our armed services. The current paper presents a review of the basic of blast physics from the near to far fields, principles of blast simulation in the laboratory, as well as recent progress in the understanding of blast induced traumatic brain injury.

8:50

2aBB3. Blast associated traumatic brain injury. David F. Moore and Michael S. Jaffee (DVBIC-HQ, WRAMC, 6900 Georgia Ave., NW, Washington, DC 20309, david.f.moore@amedd.army.mil)

Blast tissue injury is familiar problem in military medicine for gas filled organ such as the lung and gut. Solid tissue organs such as the brain have more recently come into prominence in relation to blast through the current conflicts in Iraq and Afghanistan with blast injury regarded as the "signature injury" of these wars. Enhancements in personal protective equipment allowing survivability of abdominal and thoracic injury may have exposed a previously obscured brain vulnerability. Free field explosive detonation results in rapid conversion of chemical energy into the shock wave and pressure field, the kinetic energy associated with fragments and shrapnel, thermal energy, chemical products of detonation, and electromagnetic radiation. The variation and coupling of these physical fields result in a uniquely complex problem in understanding blast biological effects especially in such a functionally intricate organ as the brain. The clinical effects of concussion following blast exposure are still undergoing evaluation most notably in the military context. Significant effort is, however, underway to determine the relative contributions of shock wave stress, the effects of cavitation, and the effects of the electromagnetic field induced by the piezo-electric effect of the skull experiencing blast associated stress in brain injury and recovery.

9:10

2aBB4. Blast induced electromagnetic pulses in the brain from bone piezolectricity. Steven G. Johnson (Dept. of Mathematics, MIT, 77 Massachusetts Ave., Cambridge, MA 02139, stevenj@math.mit.edu), K. Y. Karen Lee, Michelle K. Nyein (MIT, Cambridge, MA 02139), David F. Moore (Walter Reed Army Medical Ctr., Washington, DC), John D. Joannopoulos, Simona Socrate, and Raul Radovitsky (MIT, Cambridge, MA 02139)

The mechanisms that might lead to in-brain electromagnetic pulses from an IED-scale explosive are considered, along with whether the resulting fields might have timescales and magnitudes relevant to neurological processes. In particular, due to known piezoelectric properties of bone, it is possible for a shock wave incident on the skull to directly induce large electric fields within the brain. Using

experimental data on the piezoelectric properties of bone combined with stresses from full-head-model blast simulations, the resulting in-brain electric fields shown to have timescales and magnitudes that exceed IEEE safety standards are comparable to procedures such as transcranial magnetic stimulation that are known to have neurological effects. Not only are such electromagnetic fields at least potentially relevant to the understanding of blast-induced traumatic brain injury but they may also lead to diagnostic tools in the form of blast dosimeters that measure blast-induced head stresses via the piezoelectric fields produced just outside the skull. [This work was supported by financial aid from the Joint Improvised Explosive Device Defeat Organization (JIEDDO) through the Army Research Office.]

9:30

2aBB5. Effects of low blast levels on the central nervous system. Annette Säljö, Berndt Svensson, Maria Mayorga, Hayde Bolouri, and Anders Hamberger (Dept. of Medical Chemistry and Cell Biology, Inst. of Biomedicine, Sahlgren Acad., Univ. of Gothenburg, SE 405 30 Gothenburg, Sweden, annette.saljo@gu.se)

Anaesthetized swine in crew positions were exposed to weapons in air or to explosives underwater. Blast parameters were correlated with those in the brain. The peak pressure in the brain (*P*max brain/air) was 0.7 for a bazooka (45 kPa), 0.5 for a howitzer (10 kPa), and 0.4 for a rifle (23 kPa). The brain/water *P*max for the detonation pulse of under water explosives was only 0.1, but 0.3–0.4 for the secondary pulses. The results indicate that low-frequency spectra penetrate easier into the brain. Histological examination revealed small hemorrhages in rear regions of the brain. In rats, we investigated the effect of shock tube blasts. After exposure to 10 or 30 kPa, cognitive performance (Morris Water Maze) decreased by 50%. The intracranial pressure (ICP) increased in a dose dependent fashion to reach peak levels 6 h after exposure at 10 kPa and 10 h after exposure to 30 or 60 kPa. An initial ICP elevation took place 30 min after exposure to 60 kPa, and 2 and 6 h after exposure to 30 and 10 kPa, respectively. A prophylaxis, consisting of a 2 week intake of hydrothermally fermented cereals, reduced significantly the blast effect both on ICP and cognitive performance. [The authors thanks Svante Hjer, Samba Sensors AB. The study was supported by the Swedish Armed Forces and FMV.]

Contributed Paper

9:50

2aBB6. Effects of a simulated blast pulse train on a simple neural model. Radia Wahab (radiaraian@yahoo.com), Yunbo Liu, Victor Krauthamer (U.S. Food and Drug Administration, Silver Spring, MD), Joseph McCabe, Chantal Moratz, Ryan Egan (Uniformed Services Univ. of the Health Sci., Bethesda, MD), Vesna Zderic (George Washington Univ., Washington, DC), and Matthew Myers (U.S. Food and Drug Administration, Silver Spring, MD)

In the treatment of traumatic brain injury (TBI) due to blast-wave exposure, an understanding of the mechanisms of injury is critical for proper intervention. One mechanism of particular importance is the interaction of blast waves with neurons. In order to investigate this interaction in isolation from other TBI mechanisms, blast waves were simulated on a local (approximately 1-mm) scale using high-intensity focused ultrasound (HIFU). The acoustic impulse of the HIFU blast wave was selected to approximate the impulse of an actual blast. As a first step in a chain of increasingly complex neural models, the giant axon in an earthworm was used in this study. Intact earthworms were exposed to simulated blast waves from an HIFU transducer, and axonal action potentials were stimulated and recorded. The action potential amplitude decreased continuously with increasing impulse, with the signal typically vanishing at about 8000 Pa s. The conduction velocity showed more of a threshold effect, decreasing minimally from control values until an impulse of about 4000 Pa s was reached, at which point the decline was more rapid. Results indicate that exposing isolated neurons to HIFU-simulated blasts is a promising approach for studying TBI.

10:05—10:30 Break

Invited Papers

10:30

2aBB7. Toward the non-invasive determination of cerebral perfusion pressure. Pierre D. Mourad (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, pierre@apl.washington.edu)

Brains subjected to blast from an explosive or other sources of trauma often develop intracranial hemorrhage or edema or loose their ability to autoregulate the blood flowing into the brain (cerebral autoregulation). Such damage can, in turn, lead to increases in intracranial pressure (ICP) and decreases in cerebral perfusion pressure (CPP=arterial blood pressure—ICP) that, in turn, can lead to ischemia and/or herniation, and brain death. This is not good for the patient. To assay ICP, of intrinsic value, and CPP—two variables of critical interest to the medical staff who treat and manage patients with head injury—requires a (invasive) neurosurgical procedure. Here I describe research performed over the last several years whose target is the prediction of ICP and CPP using arterial blood pressure data as well as data derived from transcranial Doppler. In particular, I will review the process by which we collected sufficient human-derived data in order to make a decent pass at predicting ICP for patients with closed traumatic brain injury (TBI) and a great effort at predicting CPP for these and a complementary group of patients.

10:50

2aBB8. Blast-induced traumatic brain injury research at Lawrence Livermore National Laboratory. William C. Moss and Michael J. King (Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94551, wmoss@llnl.gov)

Our blast-induced TBI research has computational and experimental components. Our numerical hydro-structural simulations show that non-lethal blasts can induce sufficient flexure of the skull to generate potentially damaging loads in the brain, even if no impact occurs. The possibility that this mechanism may contribute to TBI has implications for the diagnosis of soldiers and the design of protective equipment such as helmets. Our experimental work involves designing and testing blast dosimeters, which are needed to quantify the blast environment around the soldier, independent of the mechanism(s) causing TBI. One system that uses MEMS sensors incorporated into the helmet and suspension would record peak pressure, positive-phase duration, blast direction, loads directly on the

skull, and accelerations. Another system uses less sophisticated, inexpensive, small, lightweight, disposable, unpowered sensors that act as "yes-no" gauges that indicate the blast magnitude by visual inspection of the gauge. This system is a trade-off between quantity and quality of data, which may be viable, based on current DoD needs. [This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344.]

11:10

2aBB9. Experimental and numerical evaluation of the effect of shock waves on the brain. Albert I. King (Dept. of Biomedical Eng., Wayne State Univ., 818 W. Hancock, Detroit, MI 48202, king@rrb.eng.wayne.edu)

A combined experimental and numerical study was conducted to elucidate the mechanical response of a head surrogate under air shock loading. A gel-filled egg-shaped skull/brain surrogate was exposed to blast overpressure in a shock tube environment, and static pressures within the shock tube and the surrogate were recorded throughout the event. A numerical model of the shock tube was developed using the Eulerian approach and it was validated against experimental data. An arbitrary Lagrangian–Eulerian (ALE) based fluid-structure coupling algorithm was then utilized to simulate the interaction of the shock wave with the head surrogate. A comprehensive parametric study was carried out to assess the effect of several key parameters on model response. The curvature of the surface facing the shock wave significantly affected both peak positive and negative pressures. Biological experiments exposing anesthetized rats to shock waves, using the same shock tube, produced brain injury in the form glial cell activation which in turn can adversely affect the function of axons and neurons. This injury mechanism is not the same as that for blunt impacts to the head which causes direct diffuse axonal injury.

Contributed Papers

11:30

2aBB10. Numerical simulation of mechanisms of blast-induced traumatic brain injury. Jan Arild Teland (Norwegian Defence Res. Establishment, Postboks 25, 2027 Kjeller, Norway, jan.teland@ffi.no), Fredrik Arrhén (SP Tech. Res. Inst. of Sweden, Borås, Sweden), Anders Hamberger (Univ. of Gothenburg, Gothenburg, Sweden), Morten Huseby, Reza Rahimi (Norwegian Defence Res. Establishment, 2027 Kjeller, Norway), Annette Säljö (Univ. of Gothenburg, Gothenburg, Sweden), and Eirik Svinsås (Norwegian Defence Res. Establishment, 2027 Kjeller, Norway)

Blast-induced traumatic brain injury caused by road bombs has lately become a larger part of allied injuries. The same mechanisms may also be responsible for milder injuries of similar nature, resulting from training with large caliber weapons and explosives. In this paper, the blast effects from a weapon on the brain are investigated. Using the hydrocode AUTODYN, numerical simulations of shock wave propagation into the brain are performed. The shock wave is calculated from a complete numerical simulation of the weapon, including the burning gun powder gas inside the barrel, acceleration of the projectile, and the rapid gas flow out of the muzzle. An idealized head is placed in the simulation at the position of personnel firing the weapon. Here we focus on the qualitative mechanisms of the propagation of the shock wave through the skull and into the brain. The results are compared with experiments carried out on anesthetized animals. To simulate real training scenarios, pigs were placed in position of personnel and exposed to impulse noise generated from weapons. Blast parameters in the air were correlated with those in the brain.

11:45

2aBB11. Influence of skull microstructure on blast-induced pressures in the brain. Joseph A. Turner and Jinjin Liu (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588, jaturner@unl.edu)

The interaction of blast waves with the human head is complicated by many factors including the geometry, material nonlinearities, and skull microstructure. In particular, the flat bones of the skull are comprised of the outer and inner tables (cortical bone) and the diploë (trabecular bone) on the interior. This microstructure results in both scattering and absorption of incident pressure waves. A clear understanding of these effects is needed if pressure profiles within the brain are to be accurately modeled. The focus of this presentation is on finite element wave simulations that have been developed to account for this complex organization. The models are first used to examine the scattering attenuation and coherency loss resulting from the microstructure as a function of incidence angle and frequency for plane wave incident pressure profiles. These models are then extended to more realistic pressure profiles representative of blast waves. The statistics of the microstructure are shown to play a key role in the peak pressures observed within the skull. It is anticipated that this work will lead to a better understanding of role of skull microstructure on blast-induced traumatic brain injury. [Work supported by ARO.]

TUESDAY MORNING, 20 APRIL 2010 KOZMOS, 8:00 TO 11:55 A.M.

Session 2aMU

Musical Acoustics: The Contemporary Traditional Violin

George A. Bissinger, Chair
East Carolina Univ., Dept. of Physics, Greenville, NC 27858

Chair's Introduction—8:00

Invited Papers

8:05

2aMU1. From Catgut to the Violin Society of America: The past 10 years of violin acoustics research. Fan-Chia Tao (595 Smith St., Farmingdale, NY 11735)

In the late 1990s, most of the original founding members of the Catgut Acoustical Society (CAS) had passed away and Carleen Hutchins had departed to concentrate on her new violin octet family. Nevertheless, violin acoustics research continued around the world. In 2002, the CAS and the Violin Society of America (VSA) co-sponsored a violin acoustics workshop at Oberlin College to bring together the world's leading violin acoustics researchers and violin makers. This annual VSA-Oberlin Acoustics workshop has spawned numerous collaborative projects such as George Bissinger's Strad 3D project. In 2005, the CAS merged with the VSA to become the CAS Forum. Members of the CAS Forum have continued to promote violin acoustics by organizing workshops, research projects, and presentations at international meetings of violin makers.

8:30

2aMU2. A finite element approach towards understanding violin structural modes. Colin Gough (School of Phys. and Astronomy, Univ. of Birmingham, Birmingham B15 2TT, United Kingdom)

A COMSOL finite element package has been used to model the structural modes of vibration of the violin treated as a shell structure with orthotropic arched plates. Such computations enable the physical properties of the plates, ribs, and soundpost to be varied over many orders of magnitude. This provides major insight into the nature of the coupling of the top and back plates by the ribs and the role of the soundpost in coupling the radiating "breathing" modes of the violin to the weakly radiating anti-symmetric modes excited by the bowed string induced rocking of the bridge on the central island area. Examples will be shown of the influence of such factors as plate thickness/density, anisotropy, arching, rib strength, and position of soundpost on the frequencies and excited strengths of the strongly radiating "signature" modes below 1 kHz. In addition to determining the mode shapes and frequencies of the excited modes, their contributions to the radiated sound are also computed. Comaparison will be made with experimentally observed mode shapes and frequencies.

8:55

2aMU3. The Strad 3D Project: Scientists, musicians, and violinmakers study three classic violins. Samuel Zygmuntowicz (565-A Third St., Brooklyn, NY 11215)

In 2006 the first ever three dimensional (3-D) modal laser scanning of violins was performed on three Guarneri and Stradivari violins, along with acoustic scanning and subjective evaluations. CT scanning was used to determine shape and density properties and to provide a 3-D model for future finite element analyses. These studies and images have been combined with empirical observation, photography, music recordings, and traditional documentation in an interdisciplinary survey of unprecedented scope presented in a two DVD set. From my violinmaker perspective acoustic studies of the violin have had remarkably little impact on the practice of violinmaking. Similarly, traditional empirical traditions continue to be effective in guiding violinmakers, without transferring clinically useful insights to the scientific researchers. This split is reflective of divergent vocabulary, goals, and the requirements of quantitative evidence. The Oberlin Violin Acoustics Workshops has attempted to bridge this gap by bringing together researchers, engineers, musicians, and violinmakers for interactive study, with equal weight given to these separate disciplines. The Strad3D project grows out of that ongoing collaboration and is intended to generate and collect images, documentation, and data that can be viewed and utilized from multiple perspectives and to further a mutually comprehensible dialog.

9:20

2aMU4. Global modeling of the violin radiativity profile. George Bissinger (Dept. of Phys., East Carolina Univ., Greenville, NC 27858, bissingerg@ecu.edu)

The violin's radiativity (pressure/force)profile maintains a consistent shape across quality classes, arguing for a quality-independent generalized global model. In the 196–660 Hz region the lowest cavity modes $A0$ and $A1$ and the two first corpus bending modes $B1$ generate almost all the radiativity, with the $B1$ modes treated as "pumps" for $A0$ and $A1$. The $B1$ modes have nodal patterns similar to the plate bending (primarily) modes 2 and 5, suggesting a link to the violin's critical frequency. The essentially serial character of the violin's sound chain leads naturally to a simplified expression for the averaged-over-sphere radiativity profile as a product of just two filters: (1) string-to-corpus through the tuned bridge substructure "gatekeeper") filter and (2) the "egress" filter for the vibration-radiation transformation, the latter reliably parametrized by the radiation-total damping ratio FRAD. FRAD incorporates the violin critical frequency as well as top and back plate properties in a generalized form. The gatekeeper filter on the other hand is considerably more complex; present bridge models must be augmented by systematic empirical measurements to understand the effects of varying bridge rocking mode frequency or "wing" mass. Recent three-dimensional vibration measurements provide additional insight into bridge-corpus impedance effects.

9:45

2aMU5. The violin: Perceptual studies and acoustical correlates. Claudia Fritz (Institut Jean Le Rond d'Alembert, Université Pierre et Marie Curie, UMR CNRS 7190, 4 place Jussieu, 75005 Paris, France, claudia.fritz@upmc.fr)

This talk discusses the results of experiments in which performances were replayed on different "virtual violins" in order to explore the relationships between acoustical characteristics of violins and perceived qualities. Specifically, it explores perceptual observations reported by Dünnwald [based on his measurements of over 700 instruments, J. Catgut. Acoust. Soc. (1991)] by modifying the amplitude of the resonance modes over five octave bands (thereby covering the violin's entire register). When using a subset of the most distinctive verbal descriptors of violin timbre [Fritz *et al.*, Conference of Interdisplinary Musicology (2009)] to study the relationship between human perception and these acoustical modifications, we ascertained results that partially conflict with Dünnwald's observations. In addition, the study investigated the manner by which one's perception of the violin's tone quality is affected by the magnitude of a player's vibrato as well as the damping of the violin's resonant modes. Our results do not support the conclusion that liveliness results from the combination of the use of vibrato and a "peaky" violin response. The talk concludes by discussing the limits of such psychophysical studies, suggesting future directions for psycholinguistic-based research in this domain.

10:10—10:25 Break

10:25

2aMU6. Optimizing the taper-camber relationship in bows for string instruments. John E. Graebner (2 Woodland Rd., Short Hills, NJ 07078, jegraebner@yahoo.com) and Norman C. Pickering (East Hampton, NY 11937)

A violin bow's taper (graduated diameter) and camber (inward or concave pre-bending of the stick) are important tools used by expert bow makers who know by intuition and experience how to match the two parameters. We have analyzed the static forces in bows and constructed a mathematical model that reveals a simple relationship between taper and camber. For any given taper, the model

predicts a preferred camber and vice versa. A machine has been constructed for accurately measuring the diameter and camber profiles, and measurements on several dozen bows of various degrees of playability provide support for the model.

10:50

2aMU7. Development of the *Titian* Stradivari finite element model. Michael A. Pyrkosz, Charles D. Van Karsen (ME-EM, Michigan Tech. Univ., 1400 Townsend Dr., Houghton, MI 49931), and George Bissinger (East Carolina Univ., Greenville, NC 27858)

As part of the continuing effort to understand the structural and acoustic behavior of one of Stradivari's masterpieces, a finite element (FE) model of the *Titian* Stradivari (1715) has been constructed from computed tomography (CT) scans. The CT data were used to extract high-fidelity geometry and density information specific to this violin. This violin is unique in that it is the only one of Stradivari's instruments that has been measured with a full three-dimensional mobility scan over the top and back plates and ribs as well as acoustic radiativity over a sphere. Hence this solid model can be updated and correlated with this comprehensive experimental data set. The current status of this solid modeling effort will be reviewed in the presentation.

11:15

2aMU8. The violin as a statistical energy analysis network. Evan B. Davis (Brugh Davis Acoustics, 8556 Burke Ave. N., Seattle, WA 98103)

This paper applies elementary statistical energy analysis (SEA) concepts to the violin. Modern makers have been experimenting with ultra-light violins trying to solve the problem increasing the overall output while maintaining the spectral balance of the classical violin. Anecdotal experimental evidence suggests the traditional bridge is not "appropriate" for the ultra-light violins. The ultra-light design problem is used as a case study for hybrid-SEA model. SEA modeling applies a high-modal destiny, high-frequency, approach which complements the low-frequency, low-modal density, finite element approach. The focus of the SEA modeling is the top plate or belly of the violin addressing the interaction of the cross-arching, plate thickness, wood material properties, and the dynamics of the bridge. The cross-arch of the violin is seen as a critical design feature in the violin. The current approach is more properly a "hybrid method" where the belly and box volume are represented as SEA subsystems and the bridge is represented with a two mode subsystem. The hybrid-SEA analysis demonstrates that the ultra-light's bridge mass must be in the same bridge-mass to belly-mass ratio as the traditional violin to maintain the spectral balance of the classical violin.

Contributed Paper

11:40

2aMU9. Computed cavity-air modes of Le Gruere violin and coupling to corpus modes. C. E. Gough (School of Phys. and Astronomy, Univ. of Birmingham, Birmingham B152TT, United Kingdom)

The internal cavity air modes of both a normal arched violin with f-holes and Calrleen Hutchin's Le Gruyère violin with additional holes around the ribs have been computed using COMSOL MULTIDISCIPLINARY ACOUSTICS software. The calculations underpin the need to take finite size corrections into account when evaluating the acoustically important Helmholtz or A0 air resonance. The computed modal shapes and frequencies are compared with the assumptions and predictions of the Shaw two degrees of freedom network model. This has been widely used to interpret the dependence of the A0 and A1 resonant frequencies on both cavity volume and the number of additional holes opened around the ribs. Finite size effects are shown to have a marked influence on such dependencies. They are shown to result in marked deviations from predictions based on an ideal Helmholtz resonator. The coupling of the air modes to the corpus modes that excite them is considered using a simple dynamic model. The pedictions of the model are compared with those of the Shaw network model and measurements of the dependence of A0 and A1 mode frequencies on the number and placing of additional holes opened around the ribs of the Le Gruyère violin.

TUESDAY MORNING, 20 APRIL 2010 GRAND BALLROOM V, 8:00 TO 9:00 A.M.

Session 2aNCa

NOISE-CON: Plenary

Michael J. Lucas, Chair
Ingersoll-Rand, P.O. Box 1600, Davidson, NC 28036

Chair's Introduction—8:00

Invited Paper

8:05

2aNCa1. Noise and vibration phenomena in aircraft wheel and brake systems. Todd E. Rook (Aircraft Wheels and Brakes, Goodrich Corp., 101 Waco St., Troy, OH 45373, todd.rook@goodrich.com)

There is a wide variety of noise and vibration phenomena in aircraft brake systems for which must be accounted in the design process of such systems. These phenomena include such modes as whirl and squeal, the latter of which can be quite different from its counterpart in automotive systems and has likewise received much less attention in the literature than its automotive counterpart. Consequently, an overview of such phenomena with representative results from simulations and experiments will be presented to highlight the differences. Complicating matters is that brake-induced vibration often involves strong coupling with the aircraft structure, thereby necessitating a system level understanding beyond the brake itself. This aspect poses a particular problem to a brake component supplier in how to ensure favorable noise and vibration behavior for the full aircraft system, particularly early in the development cycle. To

accomplish this goal, Goodrich has developed a successful methodology combining simulation and testing to assess the noise and vibration behavior at an earlier stage. The methodology implements simulations to guide appropriate brake component tests that are more tractable for the brake supplier yet sufficiently mimic the full system environment. With the combined simulation and testing, a reduction in actual testing can be achieved while improving component noise and vibration performance.

8:45—8:50 Questions

8:50—9:00 Announcements

TUESDAY MORNING, 20 APRIL 2010 GRAND BALLROOM I/II, 9:15 A.M. TO 12:00 NOON

Session 2aNCb

NOISE-CON and Engineering Acoustics: Experimental Techniques and Instrumentation in Noise and Vibration

Jim Thompson, Chair
Bruel & Kjaer North, 6855 Commerce Rd., Canton, MI 48187

Contributed Papers

9:15

2aNCb1. Enhancing accuracy in reconstruction of vibro-acoustic responses of a complex structure using Helmholtz equation least squares based nearfield acoustical holography. Logesh Kumar Natarajan and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

In traditional nearfield acoustical holography, the acoustic pressures in the near field are measured, and the entire acoustic field including the normal component of the surface velocity are reconstructed. This type of approach may be appropriate for reconstructing the acoustic pressure field, but not for reconstructing normal surface velocity in practice because not enough nearfield information is collected to yield satisfactory reconstruction of the normal surface velocity. In this paper, we propose to supplement the acoustic pressure measurements with a few normal surface velocity measurements, which can be used as benchmarks in optimizing the reconstruction of normal surface velocity. These data are taken as input to the Helmholtz equation least squares method based NAH to reconstruct the acoustic pressure and normal velocity on the surface of a vibrating structure. The accuracy in reconstruction is examined experimentally on a baffled rectangular plate with clamped boundary conditions subject to random excitations. To validate the results, the reconstructed field acoustic pressures are compared with those measured by microphones, and the normal surface velocities were compared with the benchmark values collected by using a scanning laser vibrometer under the same condition.

9:30

2aNCb2. Planar nearfield acoustical holography in high-speed subsonic flow. Yong-Joe Kim (Dept. of Mech. Eng., Texas A&M Univ., 315 Eng./Phys. Bldg. Office Wing, College Station, TX 77843-3123), Hyu-Sang Kwon (Korea Res. Inst. of Standards and Sci.), and Seunghwan Jung (Texas A&M Univ., College Station, TX 77843-3123)

The objective here is to develop a nearfield acoustical holography (NAH) theory that includes the effects of high-speed, subsonic flow. Recently, the speed of a transportation means such as aircraft or vehicle has significantly increased, e.g., too close to the speed of sound. As a result, the NAH data measured with a microphone array fixed on an aircraft or vehicle entail significant airflow effects such as a Doppler shift in wave number domain. Here, a convective wave equation is used to define a mapping function. The mapping function is then used to derive a NAH theory that includes the Doppler shift effect by mapping the convectional NAH procedure in the wave number domain. A numerical simulation with a monopole is performed at the high-speed airflow of Mach = 0.7. The reconstructed acoustic fields obtained by applying the proposed NAH procedure to the simulation data match well with the exact fields. Through an experiment with two loudspeakers performed in a wind tunnel at the speed of Mach = 0.1, it is also shown that the proposed NAH procedure can be used to successfully reconstruct the sound fields radiated from the two loudspeakers.

9:45

2aNCb3. A method for measuring indoor noise levels from traffic with ulterior noise sources. Daniel Oldakowski (Polysonics Corp., 405 Belle Air Ln., Warrenton, VA 20186, danielo@polysonics.com)

This paper describes a measurement protocol used to determine indoor noise levels due to motor vehicle traffic traveling on Route 395/Route 295 flyover in southeast Washington, DC. Measurements were taken over a period of 24 h, in multiple locations of a 14-story high rise building, located in the vicinity of a baseball stadium. The challenge was to remotely measure the indoor noise due to traffic in the presence of unknown and unpredictable such as construction activity inside and outside the building during the day and nearby baseball field activity at night. Multiple sound level meters were employed, inside and outside the building. One-third octave band data, with a 1-min resolution, were taken for the duration of the measurement, as is typical for indoor noise measurements. Partial recordings of the sound field time history within the building were taken to document "loud" events. During postprocessing, loud events were manually inspected to determine if those events were due to traffic noise or due to other noises.

10:00

2aNCb4. Angle dependent effects for impulse noise reduction for hearing protectors. William J. Murphy, Amir Khan, and Edward L. Zechmann (CDC/NIOSH Hearing Loss Prevention Team, 4676 Columbia Parkway, MS C-27, Cincinnati, OH 45226)

The proposed U.S. Environmental Protection Agency regulation for labeling hearing protection devices (HPDs) includes an impulsive noise reduction rating . In 2009, the American National Standards Institute Subcommittee for noise approved a revised standard for measuring the impulsive insertion loss of HPDs, ANSI/ASA S12.42-2009. The exposure at the ear in response to a forward-propagating wave depends strongly on the orientation of the head with respect to the direction of propagation. Furthermore, the insertion loss varies with the peak sound pressure level. This paper reports the results of tests performed using an acoustic shock tube to produce peak impulses of approximately 160-dB peak sound pressure level. Two manikins were evaluated: the GRAS KEMAR manikin equipped with 1/2 and 1/4 in. microphone in a GRAS 711 IEC coupler and the Institute de Saint Louis manikin equipped with a Bruel & Kjaer IEC 711 coupler equipped with a

1/4 in. microphone. The manikin heads were rotated through ±90 deg relative to the direction of the oncoming wavefront and impulsive peak insertion loss was measured according to S12.42-2009. [Portions of the research were supported by U.S. EPA Interagency Agreement No. 75921973-01-0.]

10:15—10:30 Break

10:30

2aNCb5. Development of NASA Glenn Research Center auditory demonstration laboratory facility and operational capabilities. Beth Cooper (NASA Glenn Res. Ctr., 21000 Brookpark Rd., MS 49-2, Cleveland, OH 44135, beth.a.cooper@nasa.gov), Jeff G. Schmitt (ViAcoust., Austin, TX), and David A. Nelson (Nelson Acoust., Elgin, TX)

The NASA Glenn Research Center Auditory Demonstration Laboratory (ADL) is a dual purpose facility, constructed in 2007 to support hearing conservation programs across the agency. Configured as a reverberant room, the ADL is an appropriate space for evaluating the performance of personal hearing protectors, using either human subjects or a test fixture. Hearing protector evaluations are conducted using NASA REATMASTER software, developed in partnership with the National Institute for Occupational Safety and Health. This software is available free on request to qualified laboratories, which are encouraged to participate in a collaborative program to fund continued software development. The ADL can also be configured as a free-field room to support the development of auditory demonstrations, widely used for a variety of training purposes within NASA and externally. The ADL provides an environment, sound system, and audio engineering tools for presenting and developing calibrated demonstrations of various acoustical and auditory phenomena that include fundamental acoustical and concepts, noise control principles, and simulations of hearing loss. Current work at the ADL will establish the capability of making three-dimensional surround sound recordings, which will expand the scope of the laboratory's educational products into additional areas of psychoacoustics such as binaural hearing and localization.

10:45

2aNCb6. Application of microphone frequency response and windscreen insertion-loss corrections on sound power determinations. Einar Ristroph, Michael Black, and John Phillips (1301 Arrow Point Dr., Cedar Park, TX 78613, einar.ristroph@ets-lindgren.com)

Engineering-grade product noise emission testing programs are qualified, and correction factors are applied to test results, based on "known" sound power levels of a reference sound source; i.e., a sound source that has been tested and qualified in accordance with ANSI S12.5/ISO 6926. The accuracy and uncertainty of ANSI S12.5/ISO 6926 test results are then transferred to the engineering-grade testing program and are thus of crucial importance. ETS-Lindgren recently commissioned the new Acoustic Research Laboratory at headquarters in Cedar Park, TX, and is developing an ANSI S12.5/ISO 6926 reference sound source testing program. This work included evaluation of factors contributing to accuracy and uncertainty. The application of microphone frequency response corrections (obtained from outside calibration laboratory) is discussed and data are presented. In addition, the use of windscreen insertion-loss corrections is discussed and data are presented.

11:00

2aNCb7. Anechoic chamber verification for calibration of reference sound sources. Kevin Herreman (Owens Corning Corp., 2790 Columbus Rd., R#16 Granville, OH 43023, kevin.herreman@owenscorning.com)

Confidence in the capability of a testing facility to perform repeatable /reproducible product testing mandates study of each test chamber utilized by the facility. The Owens Corning Acoustic Research Center (OCARC) provides product sound power level determination for many differing product types ranging from the power generation, appliance, automotive, to medical equipment industry utilizing a fully anechoic chamber with a removable horizontal reflecting plane. Recently the OARC completed a verification study identifying the performance of the chamber relative to the requirements of test standard ISO 6926: Requirements for the performance and calibration of reference sound sources used for determination of sound power levels and ANSI S12.5: Requirements for the performance and calibration of reference sound sources. These standards provide strict guidelines for the repeatability/reproducibility of test results. This paper will review the verification testing process and present results from the study.

11:15

2aNCb8. Developing a basis for efficient railroad horn testing. John Erdreich and Joseph Keefe (Ostergaard Acoust. Assoc., 200 Executive Dr., Ste. 350, W. Orange, NJ 07052, je@acousticalconsultant.com)

In response to complaints of excess noise from residents living in the vicinity of rail grade crossings, the Federal Railroad Administration promulgated regulations mandating minimum and maximum horn sound output levels. Older fleets require testing of horns to ensure compliance with these limits. Test requirements (no reflecting surfaces within 200 ft of the horn; wind less than or equal to 12 m/ph; no precipitation) severely limit the ability of urban commuter railroads to comply with the testing, especially in northern and coastal areas. To develop an alternate test protocol, measurements of horn sound levels were carried out in a semi-anechoic chamber and outdoors to determine if a reliable transfer function could be constructed to convert the chamber measurements to the outdoor measurements. Standard deviations of the *A*-weighted chamber measurements were smaller than standard deviations of the outdoor measurements. For one fleet, the differences between chamber measurements and outdoor measurements resulted in a consistent difference of about 41 dB(A). Locomotive supply air pressure differences at the horn were not a significant factor in sound output.

11:30

2aNCb9. Measuring recreational firearm noise. Per Rasmussen (G.R.A.S. Sound & Vib. A/S, 33 Skovlytoften, 2840 Holte, Denmark, pr@gras.dk), Greg Flamme (Western Michigan Univ., Kalamazoo, MI), Michael Stewart (Central Michigan Univ., Mount Pleasant, MI), Deanna Meinke (Univ. of Northern Colorado, Greeley, CO), and James Lankford (Northern Illinois Univ., DeKalb, IL)

Recreational use of firearms in the United States is commonplace. There are 28×10^6 Americans who consider themselves hunters and 13×10^6 went hunting in 2000. Participation in the shooting sports, without the use of properly worn hearing protection, exposes the involved persons to high levels of impulsive noise which may cause hearing loss and/or tinnitus (ear ringing). The present study was initiated to gain a better understanding of the noise exposure created by contemporary firearms using state of the art instrumentation and to ultimately increase our knowledge and awareness of this unique noise hazard. The sound pressure signal created by recreational firearms as used in hunting or target practice is characterized by a high-frequency, short duration impulsive noise. This signal is perceived by the human ear as one single, loud impulse or "shot." However, when the firearm sound level is measured with microphones capable of sampling wide frequency ranges and combined with high-speed data acquisition computer systems, the impulses can be resolved into a number of different acoustic signals related to different source mechanisms.

11:45

2aNCb10. Exposure to recreational/occupational shooting range noise versus industrial impulsive noise. Hale Marlund (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065-5402, noisedoc@aol.com)

Public shooting ranges are used primarily for recreational shooting. Certain occupations require periodic certification of firearms proficiency. Single gunfire noise is loud and impulsive. However, both recreational and occupational proficiency shooters are often exposed to firearm noise from other shooters. In such cases, the shooting noise exposure can include gunfire impulses less than 1 s apart. This situation changes noise exposure from "impulsive" to "continuous" during each open-fire session. A similar situation occurs in industrial applications when the work process involves impulsive fastening activities, such as pneumatic riveting. The similarities in high-intensity noise exposure at busy shooting ranges and production line riveting include time-weighted averages, under the current OSHA regulations, that are less than 90 dB(A) but consistently exceed 140 dB(C) peak. As a result, even though firing range and workplace rules require the use of hearing pro-

tection, it is not uncommon for recreational or occupational shooters and industrial workers to be casual about the fitting and proper use of their hearing protection devices. This paper describes the similarities in daily impulsive noise exposure at the shooting range and fastening production line. It also discusses the potential dangers in both situations with ill fitting and best practice use of hearing protectors.

TUESDAY MORNING, 20 APRIL 2010 GRAND BALLROOM III/IV, 9:15 TO 11:30 A.M.

Session 2aNCc

NOISE-CON and Noise: Automotive and Powertrain Noise and Vibration

Gordon L. Ebbitt, Cochair
Carcoustics, 29770 Hudson Dr., Novi, MI 48377

Terrence Connelly, Cochair
ESI North America, 32605 W. 12 Mile Rd., Ste. 350, Farmington Hills, MI 48334-3379

Contributed Papers

9:15

2aNCc1. Transient in cab noise investigation on a light duty diesel passenger vehicle. Dhanesh Purekar (Cummins Inc., 2727 West 450 South Walesboro, Columbus, IN 47201, dhanesh.purekar@cummins.com)

A diesel engine in cab sound quality for passenger car market is scrutinized more closely than in the mid- to heavy duty diesel truck applications. This is obviously due to the increasing expectations from the customers for gasolinelike sound quality. This paper deals with a sound quality issue recently investigated on a light duty diesel engine for a passenger van application. The objectionable noise complaint occurred during the vehicle transient operating conditions and was found to be caused by the change in the pilot quantity over a very short period of time. The root cause of the noise complaint was investigated on the noise complaint vehicle as well as simultaneously on a standalone engine in the noise test cell. Several critical combustion and performance parameters were recorded for diagnosing the issue. In addition, various standard sound quality metrics were employed to differentiate the sound quality of the objectionable noise. The issue was resolved and verified by making appropriate changes to the engine calibration without affecting key requirements such as emissions and fuel economy. Finally, the findings from the experimental tests are summarized and appropriate conclusions are drawn with respect to understanding, characterizing, and resolving this transient, combustion related impulsive powertrain interior noise issue.

9:30

2aNCc2. Two-substructure, time-domain transfer path analysis of transient dynamic response of mechanical systems with nonlinear coupling. Wenwei Jiang and Teik Lim (Univ. of Cincinnati, 688 Riddle Rd., Apt. 800L, Cincinnati, OH 45220, jiangwe@mail.uc.edu)

The traditional frequency domain transfer path analysis has become a popular method to detect and diagnose vehicle NVH problems. Unfortunately, due to its reliance on frequency transfer functions, the approach is strictly effective only for linear time-invariant system and steady-state response. This limitation prevents the method from being applied to transient and/or nonlinear behavior such as clunk, shudder, and tip-in and tip-out. In this paper, a novel time domain transfer path analysis is proposed to deal with a class of non-linear transient response problems. It combines the versatility of the transfer path analysis for tracking transmission of vibratory energy between substructures and generality of the time domain analysis for treating transient response that is also nonlinear in nature. The formulation that is derived by combining the spectral-based substructure method and a discrete, piecewise convolution theory is applied to a lumped parameter system, and the results are compared with output from a direct numerical integration of the nonlinear equations of motion. Comparison results show significant promise and appear to be usable for solving real-life vehicle problems that are highly transient with moderate level of nonlinearity present.

9:45

2aNCc3. Refinement of vehicle interior noise by reduction of driveline vibrations. Okan Tandogan, Tarkan Yapici, Mert Doganli, and Caner Sevginer (Temsa R&D & Technol. Corp., Tubitak MAM Technol. Free Zone, 41470 Gebze-Kocaeli, Turkey, okan.tandogan@temsa-argetek.com.tr)

Driveline vibrations can be a significant vibration source, which require emphasis in vehicle NVH development studies. In the early design phases, it is usually not easy to predict the vehicle level NVH performance induced by the torsional irregularities of the driveline. Establishing a model with inertia, mass, and stiffness data taken from engine, transmission and driveline suppliers can be possible and useful to initiate preliminary cautions. However, it is first, a challenging job for OEMs to receive all relevant data from suppliers to create a reliable model. Second, driveline vibration-driven NVH response of a vehicle is mostly figured out in real-time testing. Therefore, it is of great importance to treat driveline vibrations as a potential source of customer concern in vehicles and should be investigated throughout the testing phase. In this study, a dissatisfying noise behavior of a coach is determined with objective data and defined by means of engine operating condition, frequency, and excitation orders. As a next step, the problem is cascaded to its possible sources by examining the exciters' inputs and the transfer paths. Finally, the important parameters in the reduction in driveline torsional vibrations and, as a consequence, the refinement of vehicle interior noise are discussed.

10:00

2aNCc4. Calculation of optimal damping placement in a vehicle interior. Craig Birkett (Daimler Trucks North America, 4747 N Channel Ave., C3B-EA Portland, OR 97217, craig.birkett@daimler.com), Poh-Soong Tang, and Dieter Featherman (Altair Eng.)

A study was performed of a heavy duty truck cab to reduce interior noise with the objective of using a minimum amount of damping material. Several complementary tools were applied to the problem of determining damping material layout. Candidate locations for damping material were first identified by summing *A*-weighted velocities over the frequency range of interest. These velocities could be additionally weighted by acoustic sensitivity. Then an automatic optimization was performed using structural inputs to determine the optimal damping treatment from the candidate damping patches given weight constraints. Results were confirmed with testing on a vehicle dyno. The study was significant because it compared various practical methods of optimizing a vehicle interior.

10:15—10:30 Break

10:30

2aNCc5. Vibration transfer path analysis on fender noise control. Paul Liang (Harley-Davidson Motor Co., 11800 W. Capitol Dr., Wauwatosa, WI 53222, paul.liang@harley-davidson.com)

The fender is one of the major noise radiators in motorcycles. The source of the fender noise comes primarily from the engine vibration through the vehicle frame. A rubber isolator between the engine and frame can effectively break the vibration path, thus reducing the fender noise. A finite element method will be used to analyze the vibration transfer paths of an engine and a boundary element method will be used to predict the radiated noise from the rear fender.

10:45

2aNCc6. Prediction of muffler insertion loss and shell noise by a hybrid finite element acoustic statistical energy analysis model. Terence Connelly (32605 W, 12 Mile Rd., Ste. 350 Farmington Hills, MI 48334-3379, tco@esi-group.com), Steve Mattson, Dave Labyak, and Jeff Pruetz (Great Lakes Sound & Vib.)

A reactive automotive style muffler was used to evaluate and experimentally validate the numerical predictions of muffler acoustic performance. A CAD model of the silencer was developed and an acoustic FE mesh created. The interior of the muffler included two sections of perforated pipe, which were included in the cavity mesh. A hybrid FE-statistical energy analysis (SEA) numerical model was created from the finite element acoustic mesh and was excited by a diffuse acoustic field at the inlet and coupled via hybrid junctions to SEA semi-infinite fluids on both the inlet and outlet. From the hybrid acoustic model different FE-SEA modeling approaches were investigated to predict the shell noise. The muffler insertion loss and shell noise was measured experimentally using a broadband acoustic source piped into a hemi-anechoic chamber. A straight pipe with simple bends to recreate the path from muffler inlet to outlet was fabricated for comparison. Muffler insertion loss was calculated by means of a simple level difference between the silencer and the straight pipe. Shell vibration and radiated noise were also measured to validate the shell noise modeling. Agreement between experimental measurement and numerical prediction was found to be reasonably accurate up to 3 KHz.

11:00

2aNCc7. Innovative quieter aspirator design for in-car temperature sensor. Niranjan G. Humbad, Dan Silaghi, and Matthew Morris (Behr America, Inc., Behr America, Inc. 2700 Daley Dr., Troy, MI 48083, niranjan.humbad@us.behrgroup.com)

Many vehicles use automatic temperature sensor mounted on dash [instrumentation panel (IP)] to sense the vehicle interior temperature for controlling the climate comfort. One type of design requires airflow over the sensor to correctly measure the interior vehicle temperature. This suction airflow over sensor is produced by various means. The present study deals with a widely used aspirator design to draw/suck interior airflow from ventury effect caused by discharging HVAC module flow. Such a design has an inherent drawback of bringing the HVAC module noise and flow noise (generated by aspirator) to an opening mounted on dash (via a connecting tube from sensor to aspirator). In one application, this caused a customer noise complaint; noise levels were higher at frequencies above 2 kHz. This required a separate muffler part between aspirator and sensor to reduce this noise. A new innovative aspirator design was developed to replace the aspirator and muffler by a single part. This one piece assembly was designed to be easy for manufacturing but still keeping the same performance targets of the two part aspirator and muffler design. Test results of the new design are compared to the current design and shows meeting both the airflow and noise targets.

11:15

2aNCc8. Passenger vehicle exhaust cool down noise measurement and sound assessment. Antoni Szatkowski and Brian Butler (Chrysler LLC, NVH Development & Eng., P.O. Box 21-8004, Auburn Hills, MI 48321-8004, as32@chrysler.com)

Passenger vehicle exhaust cool down is getting more attention as NVH vehicles quality is more and more refined. A new test procedure was developed and presented in this paper. Data acquisition, processing, and sound qualifying are presented. Four different metrics were developed that quantify the sound level and sound quality. A few examples of different exhaust systems were analyzed. Presented methodology is a great tool for different exhaust system ranking, benchmarking, comparison, and target setting.

TUESDAY MORNING, 20 APRIL 2010 GRAND BALLROOM VII/VIII, 9:15 TO 10:15 A.M.

Session 2aNCd

NOISE-CON and Physical Acoustics: Flow Noise

Dean E. Capone, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16801

Contributed Papers

9:15

2aNCd1. The noise from flow over rough surfaces with small and large roughness elements. Stewart Glegg (Dept of Ocean and Mech. Eng., Florida Atlantic Univ., Bld 36, 777 Glades Rd., Boca Raton, FL 33431, glegg@oe.fau.edu) and William Devenport (Virginia Tech)

This paper will review theoretical and experimental results for the sound radiation from turbulent boundary layer flows over rough surfaces. Two distinctly different regimes will be considered. The first is hydrodynamically smooth surfaces for which the surface roughness does not impact the boundary layer turbulence. These surfaces have been shown experimentally and theoretically to radiate enhanced sound levels due to the scattering of the hydrodynamic pressure fluctuations by the roughness. The second regime considers large roughness elements that directly impact the boundary layer flow. Numerical calculations show that the flow separates around the large roughness elements and theoretically this implies that the roughness radiates sound in proportion to the unsteady force on each element. It is still an unresolved issue as to whether the unsteady loading is caused by the distortion of the turbulent flow around the element or the unsteady loading associated with the flow separation. This paper will describe the theoretical approaches which have been used to predict roughness noise levels in each of these two regimes and the transition from one regime to the other.

9:30

2aNCd2. Low wavenumber turbulent boundary layer fluctuating wall shear stress measurements from vibration data on a cylinder in pipe flow. William Bonness, Dean Capone, and Stephen Hambric (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, wkb3@only.arl.psu.edu)

Fluctuating wall shear stress under turbulent boundary layer (TBL) excitation is studied in this experimental investigation. A cylindrical shell with

a smooth internal surface is subjected to TBL excitation from water in fully developed pipe flow at 6.1 m/s. The vibration response of the cylinder is used to inversely determine low-wavenumber TBL shear stress levels. Both the cross-flow and streamwise directions are examined using directionally uncoupled low-order cylinder modes in the circumferential and axial directions. These data address a critical gap in available literature regarding experimental low-wavenumber shear stress data. The low-wavenumber shear stress levels in both the cross-flow and streamwise directions are determined to be roughly 10 dB higher than those of normal pressure. As is the case for various models of TBL pressure, these measurements suggest that a nearly constant value for normalized shear stress at low wavenumber is valid over a broad range of frequencies. A simple wavenumber white model form established for low-wavenumber TBL surface pressure is also shown to be appropriate for shear stress.

9:45

2aNCd3. Analysis of sound measurements inside a finite length ducted rotor system. Scott Morris, Jason Tomko, and David Stephens (Univ. of Notre Dame, Notre Dame, IN 46556, s.morris@nd.edu)

The sound generated by a ducted rotor system leads to a complicated acoustic field inside the duct. The source can have both broad-band and tonal sound features. The interior Green's function will add additional complexity due to the cut-on modes and finite-length (organ pipe) effects. In this study, a simple ducted rotor was considered experimentally and analytically in order to obtain observations and predictions of these features. The results will be an important component of modeling the structural vibration in systems where the rotor operates in an elastic shell.

10:00

2aNCd4. Aero-acoustic predictions of automotive dashboard HVAC (heating, ventilating, and air-conditioning ducts). Stephane Detry (VISTEON Systémes Intrieurs Ctr., Visteon Interior System, Technique Rue Lon Duhamel, BP 87, 62440 Harnes, France, sdetry@visteon.com), Julien Manera, Yves Detandt, and Diego d'Udekem (Free Field Technologies)

The flow-induced noise generated by automotive climate control systems is today emerging as one of the main noise sources in a vehicle interior. Numerical simulation offers a good way to analyze these mechanisms and to identify the aerodynamic noise sources in an industrial context driven by permanent reduction in programs timing and development costs, implying no physical prototype of ducts before serial tooling. This paper focuses on a numerical aeroacoustic study of automotive instrument panel ducts to estimate the sound produced by the turbulent flow. The methodology is the following: the unsteady-flow field is first computed using a CFD solver—here FLUENT. Then, the acoustic finite element solver ACTRAN computes the acoustic sound sources from these time domain CFD results. The sources are finally propagated into the vehicle interior in the frequency domain. One advantage of the technique is that the CFD computations are completely separated from the acoustic computations. This allows reusing one CFD computation for many different acoustic computations. The theoretical background is presented in the first sections of this paper. Then, the accuracy of the method for real industrial cases is demonstrated by comparing the numerical results to experimental results available at VISTEON.

TUESDAY MORNING, 20 APRIL 2010 GRAND BALLROOM VII/VIII, 10:30 TO 11:45 A.M.

Session 2aNCe

NOISE-CON, Noise, and Physical Acoustics: Sound Propagation in the Atmosphere

Kai Ming Li, Chair

Purdue Univ., School of Mechanical Engineering, 140 Martin Jischke Dr., West Lafayette, IN 47906

Contributed Papers

10:30

2aNCe1. Frequency-dependent propagation characteristics in and around forests. Michelle E. Swearingen (Construction Eng. Res. Lab., USA ERDC, P.O.B. 9005, Champaign, IL 61826, michelle.e.swearingen @usace.army.mil), Donald G. Albert (Cold Regions Res. and Eng. Lab., USA ERDC, Hanover, NH 03755-1290), Michael J. White, and Patrick Guertin (Construction Eng. Res. Lab., USA ERDC, Champaign, IL 61826)

Sound propagation in and around forests is highly influenced by the unique vegetative environment. The array of scattering objects represented can be parametrized by parameters such as tree height and diameter, spatial arrangement, and areal density. The array can be in a lattice or random configuration, depending on whether the trees were intentionally planted or naturally occurring. The trees themselves can be of uniform age and size or a mixture of sizes and types. Using data from three different study areas, this presentation will explore correlations between the physical structure of the forest and frequency-dependent attenuation of impulsive sounds.

10:45

2aNCe2. Meteorological influence on highway noise barrier performance: A case study. Paul Burgé (URS Corp., 1615 Murray Canyon Rd., Ste. 1000, San Diego, CA 92108, paul_burge@urscorp.com), Jon Sytsma, and Tom Zurburg (Michigan Dept. of Transportation)

Local meteorological conditions such as cross-winds and temperature gradients have long been recognized as factors that can negatively affect noise barrier performance. However, this relationship is seldom demonstrated with empirical data and neither are prevailing meteorological conditions routinely taken into consideration in the design of highway noise barriers. In this case study, a post-construction acoustical investigation was undertaken for a barrier that was installed as part of a recent highway improvement project and where neighborhood residents persistently complained that the barrier was not providing expected noise reduction. Several days worth of noise level measurements and meteorological data were collected and analyzed. The results of the analysis indicated that, while the area residents were not actually impacted under applicable federal and state noise policy, the noise barrier performance was indeed being significantly influenced by documented meteorological conditions. Interaction with area residents also revealed a remarkable example of the wide variation in different individuals' subjective response to virtually the same noise environment.

11:00

2aNCe3. Quasi-wavelet cascade models for intermittent random media and application to wave scattering. D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03768, d.keith.wilson@usace.army.mil), Vladimir E. Ostashev (NOAA Earth System Res. Lab., Boulder, CO 80303), George H. Goedecke (New Mexico State Univ., Las Cruces, NM 88003), Soren Ott (Risoe Natl. Lab. for Sustainable Energy, Roskilde, Denmark), and Donald G. Albert (U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03768)

Terrestrial environments often possess intermittent distributions of scattering objects. Examples include atmospheric turbulence, subsurface geology, vegetation, and buildings. A quasi-wavelet (QW) cascade process model for such intermittent random media is described, and the implications for wave scattering are examined. The QW model builds the random medium from randomly positioned and oriented, wavelet-like entities, which follow prescribed distributions for number density and energy vs spatial scale. Different types of QWs, including monopole and dipole scalar fields and toroidal and poloidal vector fields, can be combined with statistically preferred orientations to create multiple field properties possessing correlated properties and anisotropy. The spatially localized nature of the QWs facilitates construction of intermittent random fields in a manner that would be extremely challenging, if not impossible, with conventional Fourier approaches. To test the QW model, we conducted a seismic propagation experiment in the vicinity of a volcanic crater in the Mojave Desert. This site was chosen for its highly inhomogeneous, intermittent distribution of basalt and sand. Propagation of impulse signals was sampled along 864 distinct paths. Statistical distributions of seismic travel times were simulated with good success using a finite-difference, time-domain method applied to a QW model for the site geology.

11:15

2aNCe4. A numerical study of impulse propagation over a hill. Santosh Parakkal, Xiao Di, and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

In preparation for a field experiment, a numerical study of impulse propagation over a hill has been carried out. The near side of the hill corresponds to downward refraction with the associated ducted propagation. The far side of the hill corresponds to upward refraction with the associated shadow zone. Thus conditions corresponding to nighttime propagation (downward refraction) and daytime propagation (upward refraction) can be studied in the same experiment. Of particular interest is the propagation of surface waves over the hill and into the shadow zone on the far side of the hill. For selected values of hill curvature and ground impedance, numerical predictions are presented and discussed in relation to the envisioned experiment. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, New Jersey]

11:30

2aNCe5. Portable loudspeaker coverage capability for in-situ outdoor performance spaces. Juan Arvelo, Shawn Johnson, and Ronald Mitnick (Appl. Phys. Lab., The Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099, juan.arvelo@jhuapl.edu)

Observations from daily experiences reveal that sound propagation in air is influenced by the ground topography, atmospheric stratification, winds, and turbulences. A ray-based outdoor loudspeaker coverage model was developed that accounts for terrain and atmospheric conditions. Loudspeaker coverage may be adjusted to various levels of sound or speech intelligibility. Comparison against benchmark closed-form solutions and wave-based approaches demonstrate the accuracy and computational efficiency of this Gaussian-ray bundle model at aural frequencies. This model is well suited as an acoustic design aid for outdoor performance spaces. This software was packaged into an ultra-mini notebook computer with an integrated GPS antenna and a 90-m resolution worldwide terrain database to account for *in-situ* terrain effects on sound propagation, a built-in microphone for calibrated measurement of the background noise level and the loudspeaker source level, and a USB real-time logger of temperature and relative humidity. The model is also capable of importing atmospheric measurement from balloon-launched radiosondes and various atmospheric models to accurately account for atmospheric stratification in temperature, pressure, relative humidity, wind speed, and direction in forensic investigations. Atmospheric turbulence has been shown to be another important factor that must be included in the next generation portable loudspeaker coverage capability.

TUESDAY MORNING, 20 APRIL 2010 DOVER C, 8:15 A.M. TO 12:00 NOON

Session 2aNSa

Noise, Architectural Acoustics, and INCE: Ventilation, Fan, and Duct Noise Control I

Lixi Huang, Cochair
The Univ. of Hong Kong, Dept. of Mechanical Engineering, Pokfulam Rd., Hong Kong, P.R. China

Kirill V. Horoshenkov, Cochair
Univ. of Bradford, School of Engineering, Great Horton Rd., Bradford, BD7 1DP, UK

Jian Kang, Cochair
Univ. of Sheffield, School of Architecture, Western Bank, Sheffield, S10 2TN, UK

Chair's Introduction—8:15

Invited Papers

8:20

2aNSa1. Modeling the acoustic interaction between components in ventilation ductwork. Ray Kirby (Mech. Eng., School of Eng. and Design, Brunel Univ., Uxbridge, Middlesex UB8 3PH, United Kingdom)

The modeling of sound propagation in ventilation systems normally focuses on understanding the acoustic performance of individual components within the system. For example, theoretical models have long been available for ventilation silencers, elbows, branches, and bends. However, to fully understand the overall acoustic performance of a system, it is necessary also to investigate the interaction between these components. This may be achieved by using numerical methods such as the finite element method and in principle current commercial packages may now be used to examine multiple components in ventilation systems. However, commercial packages normally rely on meshing the entire ventilation system and the computational cost of such an approach quickly becomes

prohibitive at higher frequencies. An alternative, and computationally more efficient, approach is to use a modal expansion for the sound pressure field in uniform duct sections and to apply a full finite element analysis only to non-uniform elements, such a corners and/or branches. Accordingly, predictions for a ventilation system that contains multiple components are reviewed here, and the acoustic interaction between these components is examined in order to gain a better understanding of the *in situ* acoustic performance of important components such as the ventilation silencer.

8:40

2aNSa2. Simulation of acoustic noise in duct systems. M.G. Prasad and B. Rajavel (Dept. of Mech. Eng., Noise and Vib. Control Lab., Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad@stevens.edu)

Acoustical noise studies of automotive muffler, HVAC, and other duct systems are important in their design and performance. A methodology is developed for simulating the acoustical noise from a duct system which can be modeled in terms of source, duct element, and termination. In this methodology first the transfer function of the preliminary design of the duct system is obtained using system geometry, source, and radiation impedances. Then the impulse response of the obtained transfer function from the preliminary design is convolved with source signal to obtain the acoustic noise output of the system in frequency domain. In order to verify this methodology, the simulated noise is compared with experimentally measured noise for two types of duct systems, namely, a straight pipe and a simple expansion chamber. The results show good agreement between the simulated and measured noise spectra. The proposed methodology is then applied to obtain a simulation of noise reduced due to automotive muffler system. The methodology presented in this work provides the capability of simulating the noise of a duct and muffler system from its design stage before it is actually built. Further, this work could also be used for sound quality studies of duct systems.

9:00

2aNSa3. Acoustic propagation in three-dimensional, rectangular ducts with flexible walls. Jane B. Lawrie (Dept. of Mathematics, Brunel Univ., Uxbridge, UB8 3PH, United Kingdom)

Analytic expressions provide valuable benchmarks against which to check sophisticated finite element codes. In this talk some such expressions are presented for two duct configurations. Consideration is given first to the propagation of sound in a three-dimensional, unlined duct formed by three rigid walls, lying along $y - 0$, $-b \le z \le b$ and $z - \pm b$, $0 \le y \le a$ of a Cartesian frame reference, and closed by a thin elastic plate lying along $y - a$, $-b \le z \le b$. On assuming harmonic time dependence $e^{-i\omega t}$, the fluid-coupled structural waves are expressed in the form $B_n\psi_n(y,z)e^{\pm is_n x}$, $n=0,1,2,\ldots$, where $\psi_n(y,z)$ are an infinite set of non-separable eigenfunctions, s_n are the admissible wavenumbers, and B_n are the wave amplitudes. An exact, closed form expression for the eigenfunctions is presented. Various properties of the eigensystem are discussed. Second, the effect of incorporating a porous lining into the above model is considered. It is shown how the analysis is extended to include this modification. Numerical results are presented for both situations.

9:20

2aNSa4. Closed form calculation of reverberant sound propagation in a thin duct with flexible walls. Michael Panza (Dept. of Mech. Eng., Gannon Univ., 109 University Square, Eire, PA, 16541, panza@gannon.edu)

An analytical method for calculating the propagation of reverberant sound between two parallel perfectly reflecting planes in a thin duct with flexible side walls is presented. The flexible walls are modeled with a complex wave number to account for dissipation within the wall material. A mathematical model based on the Euler–Maclaurin sum formula provides an approximate closed form Green's function for the acoustic space. The method gives a set of two coupled integral equations in acoustic pressure and side wall displacement consisting of convolutions with respect to the spatial dimension along the duct. Laplace transforms for the spatial dimension are applied to the integral equations to solve for the acoustic pressure in terms of the spatial Laplace transforms of the Green functions for the reverberant acoustic space and the mechanical wave in the side wall material. A closed form solution is obtained by considering the first term of a binomial series in the Laplace domain and applying a partial fraction expansion leading to the solution in the spatial dimension along the duct. Numerical simulations show the behavior of the reverberant noise reduction provided by the flexible side walls along the duct.

9:40

2aNSa5. Exploration of a hybrid noise control system in a cylindrical duct. Gee-Pinn Too (Dept. of Systems and Naval Mechatronic Eng., Natl. Cheng Kung Univ., Tainan, Taiwan 701) and Shao-Rong Chen (Natl. Kaohsiung Marine Univ., Taiwan)

The purpose of this study is to explore the effects of sound elimination in a cylindrical duct by combining a reactive muffler and active noise control (ANC) system. Besides the exploration via experiment of the combined noise control system, predictions of desired signals in ANC system are proposed for this hybrid system. These predictions are Grey prediction based on Grey theory and signal processing for path impulse response function. In the experiment, the effects of sound elimination (such as transmission loss and insertion loss) are compared between cases with ANC systems installed before the muffler and after the muffler. The results indicate several conclusions that (1) the sequence of arrangement of muffler can influence the results of active noise control, (2) the effect of noise reduction in ANC system is influenced extremely by reference signal received, (3)the hybrid system has the advantages over a traditional muffler when the muffler is not designed for the frequency of the noise, and (4) predictions of the desired signals such as Grey prediction or path impulse response function could give a better control for the hybrid system.

10:00—10:25 Break

10:25

2aNSa6. Optimal microphone placement for active noise control in a forced-air cooling system. Raymond de Callafon and Charles Kinney (Dept. of Mech. and Aerosp. Eng., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0411, callafon@ucsd.edu)

In active noise cancellation systems with relatively small acoustic coupling, feedforward compensation is an effective methodology to create a controlled emission for sound attenuation. Especially for small electronic systems where forced air-cooling is required to control the temperature of large power sensitive components in the system, an active noise cancellation (ANC) system is a viable solution to reduce acoustic emissions. In this paper we discuss the placement of the noise source microphone for feedforward based active noise control in a forced-air cooling system. Noise source microphone placement is directed by the ANC performance of an on-line output-error based affine optimization of a linearly parametrized generalized finite impulse response filter for sound compensation. For the computation of the optimal filter, generalized or orthogonal FIR models are used as they exhibit the same linear parametrization as a standard FIR filter. The procedure is demonstrated on a small portable NEC LT170 data projector. The data projector is equipped with a shielded internal directional pick-up microphone to measure the sound created by the forced-air cooling of the projector's light bulb. Non-invasive small directional speakers located at the inlet and outlet grill of the data projector are used to minimize acoustic coupling.

10:45

2aNSa7. Reflection and transmission across partial blockages in fluid-filled flexible, non-thin-walled pipes. Iain D. J. Dupere and Wenbo Duan (School of Mech., Aerosp. and Civil Eng., Univ. of Manchester, Manchester M60 1QD, United Kingdom)

A model is presented for propagation along a flexible pipe whose thickness is not small in comparison with its diameter across a partial blockage with varying sizes and material properties. Comparison is made with Flugge's well-known thin-shell theory for a propagation along a pipe where it is found that the additional computational complexity found in the current model becomes necessary when the thickness of the shell exceeds 10% of the pipe radius. Comparison is also made with experiment both for the propagation characteristics of the pipe and for the reflection from a partial blockage. Two reflection models are presented: a crude area change model with compensation for the mass of the blockage and a more accurate model using high-order modes and matching to a flexible blockage using co-location. Reasonable agreement is found for both with the more accurate model, giving better agreement but at the expense of computational efficiency. The work is useful both for blockage detection and for detecting stenosis in blood vessels.

11:05

2aNSa8. A cochlear analog bio-mimetic muffler. Sripriya Ramamoorthy (Dept. of Otolaryngol./Head and Neck Surgery, Oregon Res. Ctr., NRC 04, Oregon Health and Sci. Univ., 3181 SW Sam Jackson Park Rd., Portland, OR 97239, ramamoor@ohsu.edu)

A noise control device, the structural acoustic silencer, is designed by adopting the tailored structural acoustic filtering exhibited by the mammalian cochlea. In the cochlea, a flexible plate of gradually varying width and thickness separates the upper and lower fluid-filled ducts. Tailoring of the flexible plate properties enables frequency-specific localization of traveling waves, which are slowed down significantly at and near the site of resonance. While such slowing of the traveling wave leads to efficient energy coupling to the flexible plate thereby reducing transmitted acoustic fluctuations, the gradually varying impedance reduces reflections and allows the frequency-component to travel to its resonance site. The design of the bio-mimetic muffler employing non-biological materials is performed using three dimensional finite element analysis and validated against experimental data. The relation between coupled dispersion and transmission loss in the noise control device is explored. The coupled wave propagation in the engineered device is compared with the wave propagation in a passive cochlea.

11:25

2aNSa9. Shaped optimization of multi-chamber mufflers with open-ended perforated inlets using a simulated annealing method. Ying-Chun Chang (Dept. of Mech. Eng., Tatung Univ., Taipei, Republic of China, ycchang@ttu.edu.tw) and Min-Chie Chiu (Chungchou Inst. of Technol., Yuanlin, Changhua 51003, Taiwan, Republic of China)

Recently, research on new techniques of single-chamber mufflers equipped with a non-perforated intruding tube has been addressed; however, the research work on multi-chamber mufflers conjugated with open-ended perforated intruding inlet tubes which may dramatically increase the acoustical performance has been neglected. Therefore, the main purpose of this paper is not only to analyze the sound transmission loss (STL) of a multi-chamber open-ended perforated inlet-tube muffler but also to optimize the best design shape within a limited space. In this paper, the four-pole system matrix for evaluating the acoustic performance is derived by using a decoupled numerical method. Moreover, a simulated annealing method has been used during the optimization process. To appreciate the acoustical ability of the open-ended perforated intruding inlet-tube and chambers inside a muffler, two kinds of traditional multi-chamber mufflers hybridized with non-perforated intruding inlet tubes (one-chamber and two-chamber mufflers) have been assessed and compared. Results reveal that the maximal STL is precisely located at the desired tone. In addition, the acoustical performance of mufflers conjugated with perforated intruding inlet tubes is superior to traditional mufflers. Also, it has been shown that the acoustic performance for both pure tone and broadband noise will increase if the muffler has more chambers.

Contributed Paper

11:45

2aNSa10. Numerical and experimental investigation of sound transmission of a tee-junction in a rectangular duct at higher-order modes. Siu-Kit Lau (Architectural Eng. Program, 203C Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681) and Kwan-Hao Leung (The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong)

Sound transmission and scattering properties in higher-order modes across the tee-junction of a rectangular duct used in ventilation and air-conditioning system were investigated numerically and experimentally. High-sound transmission of the fundamental mode and higher-order modes across the main duct is observed at eigen-frequencies of the main duct. The resonance of branch modes is suppressed by the weak modal coupling of the branch-modes and the traveling wave in the main duct at or very close to the eigen-frequencies of the sidebranch, which results in high-sound transmission of the fundamental mode and higher-order modes across the main duct and excitation of the branch modes at higher frequencies. Increases in sound scattering into higher-order mode are found when the non-planar or longitudinal branch-mode excited. In the case of co-excitation of the longitudinal branch-mode and non-planar branch-modes, a broader band-stop action in sound transmission has been observed. The results of numerical simulations were verified by experiments. A formulation of a transmission matrix based on the transfer function and a two-microphone method was shown. [The work described in this paper was supported by a grant from the Hong Kong Polytechnic University (Project A/C Code: G-U362).]

TUESDAY MORNING, 20 APRIL 2010 LAUREL A/B, 8:15 A.M. TO 12:00 NOON

Session 2aNSb

Noise, Animal Bioacoustics, and INCE: Effect of Noise on Humans and Non-Human Animals I

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Einsteinufer 25, Secr TA 7, Berlin, 10587, Germany

Chair's Introduction—8:15

Invited Papers

8:20

2aNSb1. Soundscape research in networking across countries: COST Action TD0804. Brigitte Schulte-Fortkamp (Inst. of Fluid Mech. and Eng. Acoust., Technische Universitt Berlin, 10587 Berlin, Germany) and Jian Kang (Univ. of Sheffield Western Bank, Sheffield S10 2TN, United Kingdom)

Soundscape, different from noise control engineering, is about relationships between human beings, sound environments, and society. Research in soundscape covers science, engineering, social science, humanity, and art, and it is related to many disciplines including among others acoustics, anthropology, architecture, ecology, landscape, noise control engineering, psychology, sociology, and technology. Soundscape research represents a paradigm shift in the field of environmental noise in that it combines physical, social, and psychological approaches and considers environmental sounds as a "resource" rather than a "waste." The COST Action has built a vibrant international network and delivers the foundations for soundscape indicators, an integrated database of experimental and field data, publications, and tools to support design and decision making. Therefore, the paper will discuss the underpinning science and practical guidance that supports the measurement tools and their implementation in "Soundscapes."

8:40

2aNSb2. A subject centered approach to environmental noise effects: revisiting old concepts and proposing new methods. Caroline Cance (INCAS3, Dr. Nassaulaan 9, P.O. Box 797, Assen, 9400 AT, The Netherlands, ccance@gmail.com) and Danièle Dubois (CNRS & Univ. Paris 6, Paris, 75015, France, daniele.dubois@upmc.fr)

Considering the effects of noise implies that "somebody" is affected by noise. If the noise itself is not problematic regarding its measurement (by physics), the question is less clear regarding WHO is concerned by noise? Humans, animals? Asking such a question entails some other ones: Is any living system similarly affected by noise? Generally as an organism? Or differentially as "subjects"? What scientific domains are concerned with such studies? We would address these questions from our experience and knowledge on soundscapes. Starting from psychophysics that remains within the "endemic antinomy of a world-less subject confronting a thought-less object: the antique dualism of mind and matter" [Shalins (1976)], we will consider the diversity of subjects (animal as well as human "experts" of different types, acousticians, urban planners, politicians, and inhabitants) to propose an alternative conception we name "semiophysics": it leads to reconsider the concepts of information versus meaning, as well as from a methodological point of view, the concepts of affordance versus Umwelt. Coupling field research and experimental work by accounting for meanings as relationships given to the world by the different subjects calls then for an ecological validity of laboratory investigations.

9:00

2aNSb3. Effect of combined noise sources on cognitive performance and perceived disturbance. Jin Yong Jeon and Pyoung Jik Lee (Dept. of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

An experiment was conducted to investigate the differences in cognitive performance and perceptions under road traffic noise and construction noise combined with road traffic noise. Under the conditions with individual noise and combined noise sources, an episodic memory task and semantic task were carried out in the laboratory. The subjects were asked to recall the presented words after the exposure to noises (episodic memory task) and to select the target word when the five words, including the target word and other words, were presented simultaneously (semantic memory task). Subjects also rated perceived annoyance and disturbance to the noise exposure during the experiment. The result showed that the percentage of correct answer significantly decreased with an increase in construction noise in an episodic memory task. In contrast, the semantic memory task was not impaired by the level of construction noise level. This indicates that only the retrieval task with a process of generating the internal cues was affected by the level of combined noise sources. And it was also found that the perceptions of combined noise sources were highly correlated with the result of the episodic memory task.

9:20

2aNSb4. Standardizing the measurement of natural and urban soundscapes. Robert Kull (Parsons, 7447 Central Business Park Dr., Ste. 100, Norfolk, VA 23513)

For the purpose of discussion, a simple model was presented to illustrate the contributions and interactions of various elements within an environment that make up soundscapes [Kull (2006)]. These elements should be considered when attempting to characterize a soundscape using acoustic measurements. For this presentation, the model will be used with the objective of creating a plan for taking sound measurements.

9:40

2aNSb5. Evaluating the prevalence of masking as a causal factor in wildlife responses to noise. Jesse Barber (Dept. of Fish, Wildlife and Conservation Biology, Colorado State Univ., 1474 Campus Delivery, Fort Collins, CO 80523, barber.jesse@gmail.com) and Kurt M. Fristrup (Natl. Park Service, Fort Collins, CO 80525)

Many protected natural areas are chronically exposed to noise. Noise exposure grows faster than the human populations whose activities generate noise. Data accumulate regarding masked hearing performance in animals, which can be coupled with models of sound propagation to predict reductions in the spatial extent of auditory awareness with elevated background sound levels. The emergence of predictive models of noise masking effects recommends a reassessment of field studies of wildlife responses to noise to identify the potential scope of this problem. A review of this literature reveals a substantial and diverse collection of scientific papers whose findings are plausibly related to masking effects and an increasing number of more decisive results from studies that were designed to control for other confounding factors.

10:00—10:10 Break

10:10

2aNSb6. Marine mammals and anthropogenic noise: Challenges from a management and regulatory prospective. Amy R. Scholik-Schlomer, Shane Guan, Jolie Harrison, and Craig Johnson (Natl. Marine Fisheries Service, Office of Protected Resources, 1315 East-West Hwy., Silver Spring, MD 20910)

Responsibilities of the National Marine Fisheries Service (NMFS) include conserving and recovering marine species protected under the U.S. Marine Mammal Protection and Endangered Species Acts. One of our primary objectives is to assess the risks anthropogenic noise in marine/coastal environments poses to animals in those environments and implement appropriate measures to reduce these risks. Many challenges to achieving these goals exist from both a scientific and a regulatory perspective. Accounting for the inherent complexity of source characteristics, noise propagation through the environment, and temporal/spatial overlap between sources and protected species, as well as understanding how noise exposure affects species are often quite difficult. Exposures typically are either high-level, short-term (e.g., seismic survey), or lower-level, long-term (e.g., construction project), with each presenting different risks. Establishing appropriate metrics for describing noise sources, assessing effects on individuals and on populations/stocks, as our statutes require, and ensuring the practicality of applying these metrics to real-world situations are essential. There are also often considerable data gaps, which require us to draw upon the knowledge gained from human and other terrestrial species. NMFS is currently re-evaluating and updating our acoustic criteria, which are used within our impact assessments, to reflect the best-available science on these issues.

10:30

2aNSb7. Predicting acoustic impact: Considering individuals versus populations. Brandon L. Southall (Southall Environ. Assoc., Inc., 911 Ctr. St., Ste. B, Santa Cruz, CA 95060), Adam S. Frankel, and William T. Ellison (Marine Acoust., Inc., Middletown, RI 02842, adam.frankel@marineacoustics.com)

Predictive models to assess the potential impacts of anthropogenic activities are increasing being used to assess environmental impacts. The most accurate models are, arguably, individual-based models that integrate species-specific biological and area-specific environmental data such as the acoustic impact model. These models simulate the movement of animals in four dimensions and typically include responses to both acoustic and/or non-acoustic environmental variables. The predicted exposures of individuals to sound sources may be assessed with a risk function; however, a recent meta-analysis of behavioral response and acoustic dose [Southall *et al.* (2007)] did not find meaningful linear relationships that were generally applicable between low to medium received sound levels and behavioral response. There appears to be significant variability in both species-specific and individual animal responses that are mediated by various contextual factors. Since regulatory requirements are often focused on population impact, the simplest approach to considering individual variability in behavioral response is that averaged over the population. Therefore, a mathematically linear dose-response function may be most applicable, at least within species for certain activity patterns, even if it is not supported for individual animals in all conditions.

10:50

2aNSb8. Harbor seals respond with aversion to 69-kHz pings: Implications for weighting procedures for marine mammal noise metrics. Ann E. Bowles, Stephanie K. Graves, Michael Shane (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.org), and Samuel L. Denes (Penn State Univ., State College, PA 16802)

Author Shane tracked cultured white seabass (*Atractoscion nobilis*) instrumented with 69-kHz ultrasonic coded transmitters (UCTs) in the vicinity of harbor seals (*Phoca vitulina*), later finding the bones of fish associated with UCTs. This led the authors to suspect that seals had targeted and eaten instrumented fish. To determine whether seals could detect pings, four harbor seals and a ringed seal at SeaWorld San Diego were exposed to pings from two 69-kHz and one 83-kHz UCTs and their spontaneous responses observed. The seals were not expected to respond strongly because most of the energy in the pings was close to the upper limit of hearing, but three of the four harbor seals reacted with aversion to the 69-kHz pinger with the highest source level (147 dB *re* 1 μPa), swimming into a refuge pool or jumping out of the water. The received level at the closest point of approach was estimated at 20 dB above sensation level or less. The results suggest that seals may be especially responsive to high-frequency tonal signals, and that broadband weighting functions may not consistently yield efficient exposure metrics. [Funded by NOAA; in-kind support from SeaWorld San Diego.]

11:10

2aNSb9. Modeling the exposure of greater sage-grouse to noise from industrial gas drilling rigs. Stacie L. Hooper, Sean Hanser, and Gail Patricelli (Dept. Evolution and Ecology, Univ. of Calif., Davis, 1 Shields Ave., Davis, CA 95616, slhooper@ucdavis.edu)

Natural gas and methane extraction is a growing industry in Wyoming, and some greater sage-grouse leks appear to be declining in areas near industrial sites. The goal of this project is to develop a model for understanding whether industrial noise has played a significant role in these reductions in lek attendance. A software package called NMSIM, previously developed by Wylie Laboratories to measure noise exposure from aircraft for the National Park Service, is being used. NMSIM utilizes amplitude measurements, recorded a set distance from the noise source, topographic map data, and measurements of other factors affecting sound propagation such as temperature and humidity, to build a spatially explicit model simulating how noise from the industrial sites propagates over the surrounding terrain. Simulation results are then verified using a set of noise exposure measurements taken from known locations around the gas drilling rigs. In addition to explaining historic lek attendance patterns, this model will also be used to predict how noise from new industrial sites will impact nearby greater sage-grouse leks.

11:30—12:00 Panel Discussion

TUESDAY MORNING, 20 APRIL 2010 HERON, 7:55 A.M. TO 12:00 NOON

Session 2aNSc

Noise and Architectural Acoustics: Healthcare Acoustics/Noise and Occupant Perception and Performance

Erica Ryherd, Cochair
Georgia Institute of Technology, Dept. of Mechanical Engineering, 771 Ferst Dr., Atlanta, GA 30332-0405

Kenneth P. Roy, Cochair
Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604

Mandy Kachur, Cochair
Acoustics By Design, Inc., 303 Detroit St., Ste. 304, Ann Arbor, MI 48104

Chair's Introduction—7:55

Invited Papers

8:00

2aNSc1. Evidence-based medicine meets evidence-based design: An interdisciplinary study of acoustics and sleep disruption as a context for improving teamwork. Jo M. Solet (Div. of Sleep Medicine, Harvard Med. School, 15 Berkeley St., Cambridge, MA 02138, joanne_solet@hms.harvard.edu)

Hospital noise-induced sleep arousal probability curves were derived by exposing laboratory monitored sleeping human subjects to 14 common stimuli selected and scaled from a recorded inpatient hospital sound-scape. The research process in this recently published successful project, which informed the 2010 Guidelines for the Construction of Health Care Facilities, serves as a context for discussion of the promise and pitfalls of interdisciplinary teamwork, bringing together evidence-based medicine and evidence-based design, to solve real world problems. Revealing normally unspoken professional assumptions, work ethics, record keeping requirements, and conflict of interest concerns, this program seeks to make cross field boundary breaking collaboration easier and more successful and to improve possibilities for meaningful innovation. [Work supported by AAHF, FGI, and CHD.]

8:20

2aNSc2. Hospital soundscape modeling. Selen Okcu (College of Architecture, Georgia Inst. of Technol., 247 Fourth St., Atlanta, GA 30332, selen.okcu@gatech.edu), Erica Ryherd, and Craig Zimring (Georgia Inst. of Technol., Atlanta, GA 30332)

Effective soundscape solution considerations for hospital settings can be very complex. Some acoustic qualities of these soundscapes have been shown to have potential negative impacts on occupant outcomes. To enhance these qualities, different acoustic solutions are applied in the hospital settings. Testing the effectiveness of these implications is critical but not always practical in the real settings. In this study, we examined the soundspaces of critical care settings through acoustic models. This paper will discuss the preliminary results regarding the modeled acoustic qualities of various ICU settings such as noise levels and reverb qualities.

8:40

2aNSc3. An approach to making safe and secure indoor soundscape measurements. Kenneth Good and Kenneth Roy (Armstrong World Industries, 2500 Coulumbia Ave., Lancaster, PA 17601)

Outdoor soundscape measurements are being done for many types of noise sources and locations, and the monitor equipment must be both durable and secure from both the elements and physical disturbance. Indoor measurements require a somewhat more sophisticated setup since the measurement space is often an occupied space, and in the case of hospital corridors these will likely be very active spaces 24/7. A creative yet simple approach to setting up monitor stations in hospital hallways, nurses stations, and patient rooms was developed and is being used in a number of research programs. It is imperative that valid noise measurements be made in the designated spaces without arousing undue concern or interests on the part of the occupants (both medical professionals and patients /families), and that the equipment be unobtrusive and thus secure. Monitor system mounting technique and measurement performance will be illustrated and discussed.

9:00

2aNSc4. Description and comparison of healthcare sounds. Benjamin C. Davenny and Gladys L. Unger (Acentech Inc., 33 Moulton St., Cambridge, MA 02138, bdavenny@acentech.com)

Within the clinical literature and the popular press, the noise from particular pieces of healthcare equipment or certain clinical environments is described by giving an example that is supposed to orient the reader. But these examples are often very misleading and may be in inappropriate. Outdoor everyday sounds are compared with interior healthcare sounds and steady sounds are compared with impulsive sounds. This discussion will explore better descriptions of healthcare and everyday sounds with the goal of providing better comparisons. Illustrated by examples currently in the literature we will discuss what constitutes a useful comparison and what attributes of the sounds need to be considered. Better descriptions of healthcare noise sources will help in the control of hospital soundscapes.

9:20

2aNSc5. Impact of building design on inpatient nursing unit noise levels: A case study of evidence based design. Nathan Sevener (Acoust. By Design, Inc., 4001 Murvihill Rd., Ste. 2E, Valparaiso, IN 46383)

In spring 2006, the evidence-based design process was being applied to the design of the new inpatient pavilion at Lakeland Hospital. Because of survey feedback, the hospital was aware that noise detracted from patients satisfaction with the existing facility and therefore noise level became one of the 77 metrics used in the design of the new pavilion. To help create an improved acoustical environment, the sound levels in the existing facility were monitored, the design of the new pavilion was reviewed with respect to noise control, and follow-up measurements were made after occupation. The results of that effort are presented here.

9:40

2aNSc6. Controlling patient room reverberation with a thin acoustical wall treatment. Francis J. Babineau and Amy Sparks (SoundTech Inc., 3880 SoundTech Ct., Grand Rapids, MI 49512)

A study was conducted to explore an alternate method for reducing room reverberation time in a healthcare setting. Reverberation time (RT) is known to play an important role in overall noise levels and speech privacy, both of which are key factors in patient and staff satisfaction. However, RT is difficult to control in healthcare facilities, primarily due to durability and infection control requirements. A private patient room was treated with a thin, acoustically absorptive wall treatment. The wall treatment was not a typical acoustical finish, with absorption coefficients ranging from 0.05 to 0.6. However, several walls of the room were entirely covered with the treatment to achieve reverberation control. Reverberation times were measured before and after the installation, and results were consistent with predicted values. Despite being a small study, the results suggest that a thin acoustical wall treatment that meets durability and infection control requirements can be effective for controlling reverberation in healthcare facilities and creating a more comfortable environment for patients and staff.

10:00—10:20 Break

10:20

2aNSc7. Further studies in hospital noise control at the Johns Hopkins Hospital: Part 1. Colin Barnhill, James E. West (Dept. ECE, Johns Hopkins Univ., Baltimore, MD 21218, cb@jhu.edu), Timothy Hsu, and Erica E. Ryherd (Georgia Inst. of Technol., Atlanta, GA 30332-0445)

Hospital noise levels remain well above World Health Organization guidelines. One of the main difficulties in treating hospital noise is the inability to use conventional sound absorption techniques. In intensive care units, the hospital does not allow any materials that can harbor bacteria or produce a high level of dust. This criterion prohibits conventional drop ceilings and conventional sound absorbing panels and carpet. Based on our previous panel design [M. MacLeod *et al.*, "Weinberg 5C: A case study in hospital noise control, J. Acoust. Soc. Am. 119, 3327 (2006)] a lower cost version of our hospital noise panels was implemented by replacing the wrapping with DuPont™ Tyvek©. The lowered cost makes implementation of these new panels more feasible for hospitals. In this talk, it will be shown that this new implementation produces comparable results to the original panel design. The performance of a modified DL2 measurement will be investigated as a new way to characterize hospital noise.

10:40

2aNSc8. Further studies of hospital noise control at the Johns Hopkins Hospital: Part 2. Timothy Y. Hsu, Erica E. Ryherd (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0445), James E. West, Colin L. Barnhill (Johns Hopkins Univ., Baltimore, MD 21218), Marie Swisher (Sidney Kimmel Comprehensive Care Ctr., Baltimore, MD 21231), and Natalia Levit (DuPont, Richmond, VA 23234)

It has been shown through previous research that hospital noise levels continue to rise and that there exist potential health and occupational hazards due to the soundscape. Previous studies have examined parameters such as overall average sound levels, statistical distributions of levels, and frequency content in a variety of wards. However, studying the impact of sound absorbing materials in medical facilities is still very challenging due to strict infectious control requirements. We will present a case study on installing various sound absorbing treatments in four hematological cancer wards. Four units in the Weinberg Building of the Johns Hopkins Hospital were selected due to their identical geometries, similar staff activities, and similar patient acuity levels. Researchers installed different acoustical treatments in each ward, from untreated to fully treated with prototype materials. The prototype materials tested consisted of a layer of DuPont™ Tyvek® covering a panel of fiberglass acoustic absorbing material. In each ward, extensive acoustical measurements were taken and a questionnaire survey was administered to the registered nurses on staff. Preliminary results will be presented that compare and contrast attributes such as background noise, reverberation, speech intelligibility, and subjective perception in the four different wards.

11:00

2aNSc9. Acoustic privacy and health care. Neil Moiseev (Shen Milsom & Wilke LLC, 417 Fifth Ave., New York, NY 10016)

Acoustic privacy can be differentiated into two categories: freedom from intrusive noise, such as a person snoring or wheezing in the next bed, traffic outside the windows, carts in the hallways, and footsteps on the floor above; and speech privacy—the freedom from being overheard and of overhearing others. Providing the proper acoustical environment and the protecting privacy must be a joint effort between the facility designers and hospital staff. A brief discussion of the basic requirements for speech privacy and HIPAA privacy and a quality background sound will be presented.

11:20

2aNSc10. Two healthcare buildings get the power, the noise, and the vibration from a gas turbine generator system. Chad N. Himmel (1705 W. Koenig Ln., Austin, TX 78756)

A package gas turbine generator was installed at a small, 4.3-MW cogeneration power plant of a mixed use development with office, retail, hospital, residential, and other occupancies. Shortly after commissioning, noise complaints were received from nearby medical offices about undesirable tonal noise intrusions. At another nearby hospital building, a neurosurgery suite addition would include a new intra-operative magnetic resonance imaging (iMRI) unit. Ambient vibration measurements at that site indicated prominent discrete frequencies of disturbance that could affect iMRI image quality. Investigations found loud broadband noise and dominant ground borne vibration in the vicinity of the turbine, with strong tonal peaks relating to turbine and generator rotational rates. The intrusive noise in nearby medical offices and the structure borne vibration at the hospital included matching strong tonal peaks. Various noise- and vibration-mitigating measures were proposed and some were implemented with successful results for the medical offices and new iMRI suite. Mitigating measures are discussed, along with measurement results and photographs.

11:40

2aNSc11. Strengthening the healthcare guidelines: About the new online research community. David M. Sykes (Remington Partners, 23 Buckingham St., Cambridge, MA 02138, dsykes@healthcareacoustics.org), William Cavanaugh, Gregory Tocci, and Andrew Carballeira (Cavanaugh Tocci Assoc., Sudbury, MA)

Myriad challenges and opportunities exist for researchers in the 2010 edition of the Guidelines for the Design and Construction of Healthcare Facilities. The new edition, released in January, contains the first comprehensive acoustical criteria ever included in the 60-year-old Guidelines that are now being adopted as code by most states, federal agencies, and municipalities. Enforcement—and future strengthening—of these Guidelines will require a strong, organized, well-funded research community in acoustics willing to do transdisciplinary research. Since federal agencies do not actively support research on the human impacts of noise in healthcare facilities but some private foundations do, the drafters of the Guidelines recently launched an online research community to enable research teams across the country to interact directly with interested foundations, government agencies, policy groups, and healthcare organizations. The online community has a distinguished Advisory Board of interested professionals in acoustical science, medicine, healthcare architecture, and engineering. It hosts an open database of recent research on acoustics and human health, RFPs and announcements, links to interested groups such as ICBEN, the World Health Organization and UIA-PHG, and other features such as blogs and wikis.

Session 2aPA

Physical Acoustics and Engineering Acoustics: Ultrasonics, Nonlinear Acoustics, Acousto-Optics, and Engineering Acoustics in Honor of Mack Breazeale I

Lev A. Ostrovsky, Cochair
Zel Tech/NOAA ESRL, 325 Broadway, Boulder, CO 80305-3328

Nico F. Declercq, Cochair
Georgia Tech. Lorraine, 2 rue Marconi, Metz, 57070, France

Chair's Introduction—7:55

Invited Papers

8:00

2aPA1. Mack Breazeale and E. A. Hiedemann's group at Michigan State University. David T. Blackstock (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029 and Dept. M.E., UT Austin, Austin, TX 78712-0292)

Mack Breazeale received a Ph.D. degree in physics at Michigan State University in 1957. He was one of 19 doctoral students supervised by Egon A. Hiedemann, who headed a very active group in ultrasonics at Michigan State in the 1950s and 1960s. Mack's period there, 1954–1961, is reviewed and also the history of the Hiedemann group, which was well known for its contributions to physical acoustics. A particular interest in optical measurement of ultrasonic waves led to many theses and papers in nonlinear acoustics, at that time just beginning to enjoy a renaissance. Among those who, like Mack, went on to have very active roles in the Acoustical Society of America were Laszlo Adler, Bill Cook, Logan Hargrove, and Walter Mayer.

8:20

2aPA2. Impact of Mack Breazeale on the ultrasonic studies of interfaces. Laszlo Adler (Adler Consultants Inc./Ohio State Univ., 1560 Gulf Blvd., #1002 Clearwater, FL 33767)

Over the last 50 years, Mack Breazeale made very significant contributions to Physical Acoustics. He was a great influence to his students as well as to researchers all around the world who carried on, continued, and expanded the work which was initiated by him. His work is well known on nonlinear studies of liquids and solids, acousto-optics, ultrasonic interaction at liquid-solid interfaces at the Rayleigh angle generating leaky waves, and many others. Nonlinear acoustic techniques improved material characterization of complex materials and structures significantly. The generation and detection of leaky waves opened up several applications in nondestructive evaluation. In this presentation, several of these techniques which were recently developed will be discussed. One of this is the optical detection of leaky waves at an air-solid interface which can be used to evaluate elastic properties of solids as well as surface imperfections. A nonlinear ultrasonic method to study interfaces between solid layers will be also presented. Mack Breazele leaves a legacy of knowledge, wisdom, and kindness and will always be remembered by us.

8:40

2aPA3. Acousto-optics as the background of a long-lasting friendship with Mack Breazeale. Oswald Leroy (Interdisciplinary Res. Ctr., Katholieke Universiteit Leuven Campus Kortrijk, Meensesteenweg 453, 8501 Kortrijk, Belgium)

Ever since I have known Mack Breazeale, our friendship has come along with discussions and bilateral suggestions for further research in acousto-optics. Among many other topics we had a common interest in the application of acousto-optics for nondestructive testing. For layered structures all information concerning the structure is contained in the phase information of the ultrasound investigating the material. This phase information is obtained using two or only one laser beam depending on the kind of required information. Using the technique of two laser beams the measured phase difference between the incident and the reflected ultrasonic beam is influenced by the geometrical shape of a plate, the thickness of a thin layer on a substrate, as well as by the presence of defects inside a layered material. In addition different methods are discussed to describe the diffraction of light by monofrequent, pulsed, and adjacent superposed ultrasound. In the optical nearfield of a diffracted laser beam, it is even possible to reconstruct all of the acoustic wave parameters. The ability of measuring the amplitude and phase of reflected and transmitted waves is a powerful tool to discover information about the surface on which scattering of sound occurs.

9:00

2aPA4. Substructural organization and acoustic harmonic generation in fatigued metals. John H. Cantrell (NASA Langley Res. Ctr., MS 231, Hampton, VA 23681, john.h.cantrell@nasa.gov)

Since the discovery of acoustic harmonic generation in solids by Breazeale and Thompson [Appl. Phys. Lett. **3**, 77 (1963)] and Gedroits and Krasilnikov [Sov. Phys. JETP **15**, 122 (1963)], nonlinear acoustics has gained popularity as a tool for nondestructive materials evaluation. A longstanding problem has been the assessment of fatigue damage. Harmonic generation measurements are

shown to provide for the first time a quantitative, unambiguous means to assess the state of fatigue in metals from the virgin state to fracture. The salient features of an analytical model are presented that account for the microelastic-plastic nonlinearities resulting from the interactions of an acoustic wave with self-organized dislocation substructures and cracks that evolve during cyclic loading. The model predicts a monotonic increase in the nonlinearity parameter of several hundred percent over the life of the material. Generally, the increase in the nonlinearity parameter during the first 80%–95% of fatigue life is dominated by the evolution of organized dislocation structures, while the last 5%–20% is dominated by crack growth. Applications of the model to aluminum alloy 2024-T4, 410Cb stainless steel, and IN100 nickel-base superalloy yield excellent agreement between theory and experiment.

9:20

2aPA5. A diffraction correction for the nonlinearity parameter measured by the harmonic generation technique. William T. Yost (NASA-Langley Res. Ctr., 3B E. Taylor St, Rm. 285, MS231, Hampton, VA 23681-2199, william.t.yost@nasa.gov)

Practical applications of harmonic generation to determine the nonlinearity parameter Beta generally tend toward the lower drive (fundamental) frequencies. As a result, diffraction effects on Beta become more significant. Derivations of a number of different diffraction correction formulas, which are applied for a piston source and a receiver located some distance away, are in the literature. In the paper by Blackburn and Breazeale [J. Acoust. Soc. Am. **76**, 1755–1760 (1984)] the correction formulas were applied to the amplitude of the fundamental frequency and the results on Beta were experimentally tested. By building on this earlier work, a diffraction correction formula that includes a factor from a numerical integration algorithm applied to the harmonically generated wave is provided. The formulation of this algorithm is discussed. Experimental results are given for measurements of Beta in AA 2024 to illustrate the agreement.

9:40—10:00 Break

10:00

2aPA6. Nonlinear acoustics: A transition from laboratory to a practical fatigue assessment tool. Jeong-Kwan Na (SID, Univ. of Dayton Res. Inst., 300 College Park, Dayton, OH 45469, jeong-kwan.na@wpafb.af.mil)

The ultrasonic second harmonic generation technique measuring the nonlinearity parameter has been developed and used in laboratories for several decades throughout the world. Many students and visiting scholars from numerous countries devoted their passions at the physical acoustics laboratories led by Dr. Mack Breazeale for 50 years. During that period of time, nonlinear elastic properties of single crystals, polycrystalline alloys, superconductors, ceramics, and composites were measured. New measurement systems and techniques were also developed to understand temperature dependent linear and nonlinear elastic behaviors of these materials over a temperature range from the liquid helium to Curie temperatures of piezoelectric ceramics. It took years of research and development efforts with a persistent funding before a practical fatigue damage measurement technique was finally developed. The transition from a rack full of equipment to a portable fatigue damage measurement system, a continuous iteration process was inevitable. The current system, specifically designed for steam turbine blades fatigue inspections, consists of a probe fixture and an industrial grade lunch-box computer with custom designed signal processing cards.

10:20

2aPA7. Nonlinear interaction of ultrasonic waves: A tribute to Mack Breazeale. Murray S. Korman (Dept. Phys., U.S. Naval Acad., Annapolis, MD 21402)

While at Brown University (~1977), this graduate student was fortunate enough to learn about Mack Breazeale's important early work in nonlinear ultrasonic waves in solids [J. Appl. Phys. **36**, 3486 (1965)] through his thesis advisor (R. T. Beyer), who had friendly and most collegial interactions with Mack, through the ASA. When I met Mack Breazeale at ASA Meetings, he would discuss how he made some of his measurements and helped me understand some complexities involving the scattering of sound by sound. While Mack Breazeale was Associate Editor of nonlinear acoustics he gave a wealth of good advice (over a period of several years) on how to improve a manuscript on crossed beam scattering in turbulence and divide the material into at least two parts—which was done. When the paper was out, one agreed with Mack that it was much better and worth the effort. Years later, doing summer research at NCPA, Mack helped improve a manuscript on nonlinear acoustic landmine detection. As a tribute to Mack Breazeale, his work on "Quantum mechanical theory on nonlinear interaction of ultrasonic waves" (with I. L. Bajak) [J. Acoust. Soc. Am. **68**, 1245 (1980)] will be discussed in some detail.

Contributed Papers

10:40

2aPA8. Peculiarities of acoustic beam reflection from a fluid-solid interface. Oleg A. Sapozhnikov (Appl. Phys. Lab., Univ. of Washington, and Dept. Phys., Moscow State Univ., Russia, olegs@apl.washington.edu), Alexander A. Karabutov, Jr., and Vladimir G. Mozhaev (Moscow State Univ., Russia)

Mack Breazeale and his colleagues published several papers on experimental study of non-specular reflection of ultrasonic beam from a fluid-solid interface [Breazeale *et al.*, J. Acoust. Soc. Am. **56**, 866 (1974); J. Appl. Phys. **48**, 530 (1977)]. To study fairly unusual details of the reflected beam structure, Schlieren visualization was employed, which clearly confirmed theoretical predictions. The present talk was motivated by this elegant approach of Breazeale. One of the goals was to improve reflected beam imaging by using pulsed Schlieren technique. The second goal was to observe a growing interface wave. It is known that the secular equation for acoustic waves at fluid-solid interfaces yields the common leaky wave and its complement. This complementary wave grows instead of decays with propagation and is time-reversed compared to the leaky wave. Ultrasonic pulses and their reflections were visualized using Schlieren imaging and stroboscopic flashing of a semiconductor laser. The source was a broadband (0.5–3 MHz) single element plane transducer. Reflectors were aluminum blocks with fine angular adjustments in an optically transparent water tank. The wave reflection and transmission were also studied numerically using finite differences. [Work was supported by RFBR, NIH, and NSBRI grants.]

10:55

2aPA9. Fatigue damage evaluation using harmonic generation in polycrystalline copper. Brian Reinhardt and Bernhard Tittmann (Dept. of Eng. Sci. and Mech., Penn State Univ., 212 Earth Eng. Sci. Bldg., University Park, PA 16802)

A fundamental goal of ultrasonic nondestructive evaluation is to characterize material defects before failure. During material fatigue, dislocations tend to nucleate and become sources of stress concentrations. Eventually, cracks start to form and lead to material failure. Recent research has indicated that nonlinear harmonic generation can be used to distinguish between materials of high- and low-dislocation densities. This research reports nonlinear harmonic generation measurements to distinguish between those areas of high- and low-dislocation densities in copper bars. The copper bars were subjected to flexural fatigue. Periodic scans were taken in order to track dislocation development during the fatigue life of the material. We show that this technique provides improved early detection for critical components of failure.

11:10

2aPA10. Simulation of nonlinear acoustic waves based on the semi-Lagrangian method. Glauber T. Silva and Andre Nachbin (Inst. of Phys., Univ. Fed. Alagoas, Maceio, Alagoas 57072-970, Brazil, glauber@pq.cnpq.br)

In this work, we introduce the semi-Lagrangian numerical method for one-dimensional nonlinear wave propagation in acoustics. The method is suitable to simulate interacting waves in fluid media. In the method, the acoustic fields are expressed in terms of Riemann invariants. These variables are computed along the characteristic curves one step back in time. The unknown characteristics are obtained by the Runge–Kutta method. The Riemann invariants are evaluated in the grid points using cubic interpolation. The method is validated comparing the numerical solution to the Earnshaw solution for simple waves. Results show that the proposed method excels the Courant–Frederich–Lewy condition at least fourfold. The total numerical error is 12% and 3% using 16 and 32 points per wavelength, respectively. The proposed method is applied to simulate the interaction between a finite-amplitude tone with a Gaussian noise. Solutions are in qualitative agreement with earlier studies on tone and noise interaction [D. A. Webster and D. T. Blackstock, J. Acoust. Soc. Am. **64**, 687–693 (1978)]. [Work partially supported by CNPq.]

11:25

2aPA11. Ultrasonics in solids with frequency-dependent nonlinearity parameters. Paul A. Elmore (Naval Res. Lab., Marine Geosciences Div. Stennis Space Ctr., MS 39529, paul.elmore@nrlssc.navy.mil)

This paper reviews the current state of modeling the nonlinear generation of second-harmonic ultrasonic waves in heterogeneous solids. Experiments measure the rate in which the higher-order harmonics grow with wavenumber, k, and propagation distance, a. The lowest-order nonlinearity parameter, quantifying the growth rate of the second-harmonic amplitude, A_2, experimentally is $\beta \propto \lim A1 \rightarrow 0(A1/A22k2a)$, where A_1 is the fundamental harmonic amplitude. In crystals and amalgamated metals, β is constant, and the corresponding equation of motion can be written with one nonlinear term to a good approximation. In heterogeneous materials with granular contacts, however, such as lead-zirconate-titanate (PZT) and graphite-epoxy composites, β changes with wavenumber, and the corresponding equation of motion requires more nonlinear terms. An equation of motion updated from the one for crystals does provide predictions of β that appear to match the experimental results; however, it is inadequate for providing velocity predictions in graphite-epoxy composites, where there is significant dispersion, and does not explicitly account for the hysteric behavior of grain contacts. A recent model that accounts for grain contacts asserts that it can account for the frequency dependent β. This paper will review these models and compare predictions for β with PZT and composites data.

TUESDAY MORNING, 20 APRIL 2010 ESSEX A/B/C, 9:00 A.M. TO 12:00 NOON

Session 2aPP

Psychological and Physiological Acoustics: Application of Psychoacousticss to the Impaired Listener, Challenging Environments, Hearing Aids, and Cochlear Implants (Poster Session)

Sandra Gordon-Salant, Chair

Univ. of Maryland, Dept. of Hearing and Speech Science, College Park, MD 20742

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbers papers will be at their posters from 10:30 a.m. to 12:00 noon.

2aPP1. Comparison of central masking versus monotic masking in non-simultaneous masking conditions. Mahnaz Ahmadi and Lawrence L. Feth (Dept. of Speech and Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210)

The growth of non-simultaneous masking was compared in a central masking versus a monotic masking condition for the same masker and signal parameters. Detection thresholds of normal listeners were measured utilizing a 2AFC up-down adaptive-tracking procedure. Signal-to-masker time intervals were 0, 2, 5, 10, 20, and 50-ms. Brief sinusoidal signals at 0.5, 1, 2, and 4 kHz were masked by a two-octave band of noise of 200 ms duration. The masker was centered on and off the frequency of the signal. Detection thresholds were greater for signal-to-masker intervals less than 50 ms in forward masking using both central and monotic masking conditions. Backward central masking did not result in significant threshold shifts. High-frequency signals produced greater central masking effects; however, signals at lower frequencies were more effectively masked in monotic conditions. In contrast with monotic masking conditions, off-frequency maskers did not produce a clear central masking effect. Results of this study are consistent with neurophysiologic findings of medial olivocochlear bundle response characteristics and support the idea that central masking is mediated by efferent fibers in humans.

2aPP2. Comodulation masking release with signal-masker interactions represented in the envelope. Robert H. Pierzycki and Bernhard U. Seeber (MRC Inst. of Hearing Res., Univ. Park, Nottingham NG7 2RD, United Kingdom, rp@ihr.mrc.ac.uk)

In normal hearing, the threshold of a tone masked by a modulated narrow-band on-frequency masker (OFB) can be reduced if correlated modulation is present on spectrally distant flanking bands (FBs), an effect

known as comodulation masking release (CMR). Since electric hearing with current cochlear implants is based on envelope information, comodulation of envelopes on multiple electrodes might also be beneficial for detection in electric hearing. CMR was investigated with normal-hearing participants listening to unprocessed or vocoded stimuli. In Experiment 1, tone thresholds were determined when masked by a sinusoidally amplitude-modulated band of noise (SAM and OFB) and by zero to four FBs of noise whose envelopes were either co- or anti-modulated with the OFB envelope. In Experiment 2, envelopes of those signals were extracted in a vocoder and used to modulate noise or sinusoidal carriers, thereby replacing the original temporal fine structure (TFS). Significant CMR of 3–10 dB was found in unprocessed conditions and although reduced to 2–6 dB, CMR was still significant after vocoding. CMR did not differ significantly between the sine and the noise vocoder, suggesting that the applied SAM determined the magnitude of CMR. Since CMR withstands vocoding, comodulation is hoped to improve detection in electric hearing.

2aPP3. The speech-critical band for vowels in steady and fluctuating backgrounds. Eric W. Healy (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, healy.66@osu.edu), Kimberlee A. Crass (Univ. of South Carolina, Columbia, SC 29208), and Frederic Apoux (The Ohio State Univ., Columbus, OH 43210)

Previous work investigating the frequency resolution employed to process English vowels indicated that the speech-critical band (S-CB) for vowels was greater than the psychophysical critical band. This result suggests that the density of information in the acoustic signal is below the resolving power of the auditory system, and therefore functional resolution is limited by the signal. In the current investigation, the S-CB for vowels was measured in steady and fluctuating backgrounds. Thirty-six normal-hearing listeners heard vowels in /hVd/ context. The stimuli were restricted in overall bandwidth to 1.5 octaves centered at 1500 Hz and presented using vocoder-like processing as 1, 3, 5, 7, 9, or 11 low-noise noise carrier bands. Recognition increased and reached asymptote as the number of constituent bands increased. The carrier bandwidth at performance asymptote was again found to be larger than the psychophysical critical band. Moreover, the resolution of the acoustic signal at performance asymptote was similar across conditions in which the stimuli were presented (i) in quiet, (ii) in steady noise, or (iii) in four-talker babble at various signal-to-noise ratios. These results suggest that the current measure of speech frequency resolution is robust across a number of adverse listening conditions. [Work supported by NIDCD.]

2aPP4. Release from speech-on-speech masking under degraded signal conditions. Virginia Best (School of Med. Sci., Univ. of Sydney, Sydney, New South Wales 2006, Australia and Hearing Res. Cntr., Boston Univ., Boston, MA 02215, ginbest@physiol.usyd.edu.au), Nicole Marrone (Northwestern Univ., Evanston, IL 60208 and Boston Univ., Boston, MA 02215), Christine Mason, and Gerald Kidd, Jr. (Boston Univ., Boston, MA 02215)

Previously, Marrone *et al.* [J. Acoust. Soc. Am. (2008)] compared spatial release from masking (SRM) for a three-talker mixture of similar sentences in normal-hearing (NH) listeners and listeners with sensorineural hearing loss (HL). The HL group showed significantly less SRM. In an earlier study [Marrone *et al.*, J. Acoust. Soc. Am. (2008)] the NH group showed less SRM when the masker sentences were time-reversed while in the colocated case a large "reversed-masker release" (RMR) was also found. To investigate these findings, some listeners from the HL group were tested in the reversed-masker conditions. The difference in SRM between listener groups was much smaller for reversed speech. However, the HL group also had much less RMR. Both the overall target-to-masker ratios (TMRs) at which comparisons are made and the spectral smearing associated with sensorineural hearing loss were likely factors, perhaps affecting amounts of energetic and informational masking. We explored these factors in a new group of NH listeners in a similar task with both forward and reversed speech systematically degraded. Performance was measured at seven TMRs for unprocessed as well as 32-, 16- or 8-channel vocoder speech. The results provide qualified support for an influence of TMR and spectral degradation on both SRM and RMR.

2aPP5. Auditory stream segregation using amplitude modulated vocoder bandpass noise. Yingjiu Nie and Peggy Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455, peggynelson@umn.edu)

We investigated the contribution of amplitude modulation (AM) rate and spectral separation to stream segregation of vocoder bandpass noises. Stimulus sequences were repeated pairs of A and B bursts, where bursts were white noise or vocoder bandpass noise carrying sinusoidal AM (100% modulation depth). Bursts differed either in the center frequency of the noise, or the AM rate, or both. Eight vocoder bands were used. The lowest four bands (1-2-3-4) were combined into one bandpass noise (B bursts) and the higher three bands (3-4-5, 4-5-6, and 6-7-8) were combined to constitute the A bursts. Results show that stream segregation ability increases with greater spectral separation. Larger AM rate separations were associated with stronger segregation abilities, but not when A and B bursts were both white noise. Significant inter-subject differences were noted. Results suggest that, while both spectral and AM rate separation separations could be cues for auditory stream segregation, stream segregation based on AM rate is more successful when combined with spectral separation. Correlations between segregation ability and understanding of vocoded speech will be discussed.

2aPP6. A test battery to assess localization ability in simulated complex acoustic environments. Stefan Kerber and Bernhard U. Seeber (MRC Inst. of Hearing Res., Nottingham NG7 2RD, United Kingdom, s.kerber@ihr.mrc.ac.uk)

In complex acoustic environments, following a target sound is made difficult by the presence of noise and reverberation and especially hearing impaired people struggle understanding speech in such settings. Localizing the speaker is important in such situations to follow discussions and to gain additional benefit from visual cues. A test battery is proposed to quantify localization performance in different realistic environments and with laboratory tests when noise and reverberation are present. The aim is to relate real-life performance to that in simplified tests. In the localization tests, participants indicate the sound location with a visual pointer. Sounds are played in an anechoic chamber from loudspeakers across the frontal hemifield. Localization performance is measured in quiet, in diffuse background noise, and with reverberation of simulated rooms. The ability to cope with a single reflection is assessed in a precedence effect paradigm where participants localize a sound and its delayed copy for various delays and levels. The test battery is completed by questionnaires (e.g., SSQ), a speech, and a cognitive test. By cross-comparing results from the tests, we attempt to predict the performance in real-world environments from outcomes of simplified tests particularly in the presence of hearing impairment.

2aPP7. On the possible influence of spectral- and temporal-envelope cues in tests of sensitivity to temporal fine structure. Christophe Micheyl (Dept. of Psych., Univ. of Minnesota, 75 East River Rd., Minneapolis, MN 55455-0344, cmicheyl@umn.edu), Huanping Dai (Univ. of Arizona, Tucson, AZ 85721), and Andrew J. Oxenham (Univ. of Minnesota, Minneapolis, MN 55455-0344)

The role of temporal fine structure (TFS) in speech and pitch perception has attracted considerable attention. Recent studies have suggested sensitivity to TFS even at high frequencies (8 kHz), well beyond the known limits of phase-locking in mammals, and a reduced ability to use TFS in hearing-impaired listeners. These conclusions were based on tasks involving the discrimination of complex tones or iterated-rippled noise, which had different power spectra but were (a) bandpass-filtered in a way that reduced tonotopic excitation-pattern (EP) cues and (b) had random component phases to eliminate temporal-envelope (TE) cues. In this study, we examine the possibility that residual EP and/or TE cues may have been available to the listeners in these experiments. Analytical and computational analyses indicated that although systematic TE differences were absent at the level of the stimuli, they were likely present at the output of the cochlea. Empirical studies tested whether the available TE and EP information might have influenced

performance. Preliminary results suggest that although listeners may not have been sensitive to the available TE cues, performance may have been influenced by EP cues in a way that could also explain the deficits shown by hearing-impaired listeners. [Work supported by NIH R01 DC05216.]

2aPP8. Discrimination of repetitive intervals by younger and older listeners. Peter Fitzgibbons (Dept. Hearing, Speech, and Lang Sci., Gallaudet Univ., 800 Florida Ave., NE, Washington, DC 20002) and Sandra Gordon-Salant (Univ. of Maryland, College Park, MD 20742)

The study measured listener sensitivity to increments in the inter-onset intervals (IOIs) separating successive 20-ms 4000-Hz tone bursts in isochronous sequences. Stimulus sequences contained from 2–6 tone bursts, separated equally by IOIs in the range of 25–100 ms across stimulus conditions. Difference limens (DLs) for increments of all tonal IOIs were measured to assess listener sensitivity to changes in sequence rate. A DL was also measured for increments of a single IOI located at a fixed position within 6-tone sequences. Listeners included younger and older normal-hearing adults and older adults with high-frequency hearing loss. The results revealed that the relative DLs for sequence rate decreased as the magnitude of the reference IOI and the number of sequence components increased. The relative DL for a single interval embedded within a six-tone sequence was smaller than corresponding DLs measured with two-tone sequences, but only for brief reference IOIs. The discrimination performance of the older listeners was poorer than that of the younger listeners, especially for two-tone sequences with the shortest reference IOIs. The findings are interpreted within the context of multiple-look mechanisms and possible age-related differences in the sensory coding of signal onsets. [Research supported by the National Institute on Aging, NIH.]

2aPP9. Age-related differences in auditory spatial attention depend on task switching complexity. Gurjit Singh (Dept. of Psych., Univ. of Toronto, 3359 Mississauga Rd. N, Mississauga, Ontorio L5L 1C6, gurjit@psych.utoronto.ca), M. Kathy Pichora-Fuller, and Bruce A. Schneider (Univ. of Toronto, Mississauga, ON, L5L 1C6, Canada)

We investigated the role of simple and complex switching of auditory attention in a multi-talker, multi-spatial listening situation with target location uncertainty. In all conditions, a target sentence from an edited version of the CRM corpus was presented from one spatial location and competing sentences from two different locations, with cues specifying the target's callsign identity and the probability of its location. Four probability specifications indicated the likelihood of the target being presented at the left, center, and right locations (0-100-0, 10-80-10, 20-60-20, and 33-33-33). In conditions requiring simple switches of attention, the task was to report key words from the target sentence. In conditions requiring complex attention switching, when target callsigns were presented from one of the unlikely locations, the listener's task was to report key words presented from the other unlikely location. A total of eight younger and eight older adults who had normal audiometric thresholds below 4 kHz participated. The key finding is that, whereas both age groups performed similarly in conditions requiring simple switches of attention, older performed worse than younger listeners in conditions requiring complex switching. Switching complexity may explain, in part, why older adults with relatively good audiograms report difficulty communicating in complex listening situations.

2aPP10. Psychophysical tuning curves and recognition of highpass and lowpass filtered speech for a person with an inverted V-shaped audiogram. Vinay Nagaraj (Dept. of Electrons and Telecommunications, Acoust. Res. Ctr., Norwegian Univ. of Sci. and Technol. (NTNU), NO-7491 Trondheim, Norway) and Brian Moore (Univ. of Cambridge, Cambridge CB2 3EB, United Kingdom)

A single subject whose audiogram resembled an inverted V shape (good hearing at 4000 Hz, poorer hearing at other frequencies) was tested. Results of the TEN(HL) test suggested that a dead region (DR) in the cochlea was present at all test frequencies from 500 to 3000 Hz, but no DR was present at 4000 Hz. Psychophysical tuning curves obtained using signal frequencies of 2000, 3000, 4000, and 6000 Hz showed upward shifted tips for the lowest two signal frequencies and a downward shifted tip for the highest frequency. The results suggested a functioning region extending from 3900 to 5100 Hz, with DRs outside that range. The identification of nonsense syllables, amplified according to the Cambridge formula, was measured as a function of lowpass or highpass filter cutoff frequency. The results suggested that useful speech information could only be extracted from a limited frequency range around 4000 Hz.

2aPP11. The relationship between quiet threshold and the forward-masking temporal effect. Elizabeth A. Strickland (estrick@purdue.edu) and Yonit A. Shames (Dept. SLHS, Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907)

Research in our laboratory has shown that the temporal effect (TE) in simultaneous masking is consistent with a decrease in gain, possibly mediated by the medial olivocochlear reflex (MOCR). The TE in simultaneous masking is a decrease in threshold signal-to-masker ratio for a signal at masker onset when a precursor is added. This work has been extended to a forward-masking paradigm. A growth of masking (GOM) function is measured with a short-duration off-frequency masker (which should not activate the MOCR) and a 4-kHz signal. Then the masker level is fixed at a point on the lower leg of the GOM function, and threshold is measured with a long-duration precursor which is intended to activate the MOCR. The estimated input-output function is compared for a precursor at and well below the signal frequency. The difference in thresholds is the TE, which is also a measure of the change in gain. In simultaneous masking, the size of the TE decreases with increasing quiet threshold. In the present study, this relationship was examined for the forward-masking TE. The TE was measured as a function of precursor level for listeners who had a range of quiet thresholds, including listeners with mild cochlear hearing impairment. [Research supported by a grant from NIH(NIDCD) R01 DC008327.]

2aPP12. Interaural time difference thresholds as a function of the duration of the beginning of a sound in hearing-impaired adults. Michael A. Akeroyd and Fiona H. Guy (MRC Inst. of Hearing Res., Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow, United Kingdom, maa@ihr.gla.ac.uk)

The precedence effect indicates that the auditory system places most weight on spatial information at the onset of a sound. In this experiment, we measured how well hearing-impaired adults can ignore what comes immediately after the onset of a sound, as any limitations in that should interfere with localization in environments with lots of reflections. The stimulus was designed to be a simplified analog of a room impulse response: it consisted of two bursts of speech-shaped noise concatenated together: the first was short and given some ITD (representing an ideal onset that marked direction correctly), but the second was far longer and was interaurally uncorrelated (representing later reflections and reverberation from all possible directions). We measured psychometric functions (over the ITD and duration of the first burst) to determine the minimum duration of the first burst needed to report its direction. Presently, 19 listeners have completed the task. No significant correlation was found between minimum first-burst duration and hearing loss (better-ear loss = 3–50 dB), but the individual variation was considerable, especially so in those listeners with larger hearing losses. The results indicate that some listeners will have difficulty in ignoring irrelevant information after a sound's onset.

2aPP13. The effects of development and hearing impairment on the ability to understand speech in temporally and spectrally modulated noise. Joseph W. Hall, III, Emily Buss, and John H. Grose (Dept. Otolaryngol., Univ. UNC Chapel Hill, CB 7070, Chapel Hill, NC 27599, jwh@med.unc.edu)

This study examines the ability of hearing-impaired children to receive speech recognition benefit from temporal modulation, spectral modulation, or combined spectral and temporal modulation of a background noise. The task involves identification of words within the context of meaningful sentences presented in a speech-shaped noise background. Control groups include normal-hearing adults and children and hearing-impaired adults. The age range of the children is approximately 5–10 years. In the procedure, the masker level is held constant and the speech level is adaptively varied to track a criterion percent correct. Preliminary data suggest that both hearing-impaired children and hearing-impaired adults show a poorer than normal

ability to benefit from temporal and spectral modulation. The hearing-impaired children require approximately 2-dB higher signal-to-noise ratio than the hearing-impaired adults for most conditions. The normal-hearing adults and children show better masked thresholds than their hearing-impaired counterparts, with normal-hearing adults generally performing 3–5 dB better than normal-hearing children across conditions.

2aPP14. Performance of phonemically targeted processing in conjunction with compression processing with spectral enhancement. Jeffrey J. DiGiovanni (Auditory Psychophysics and Signal Processing Lab., School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701, digiovan@ohio.edu), Janet C. Rutledge (Univ. of Maryland, Baltimore City, Baltimore, MD 21250), and Chessy S. Umble (Ohio Univ., Athens, OH 45701)

Sensorineural hearing loss is strongly linked to poorer speech intelligibility, especially in noise. The goal of the present study is to test, individually and in combination, two signal processing strategies designed to improve both consonant and vowel perception. In the first strategy, specific consonants were targeted for processing to increase amplitude and duration. For consonants with a duration increase, the adjacent vowel was decreased proportionately in order to maintain overall word and sentence duration. Second, Col-SE, an adaptive compression processing strategy incorporating spectral enhancement, was used to process stimuli. Hearing-in-noise-test sentences were presented monaurally to normal-hearing and hearing-impaired adults through an insert earphone in the presence of speech-shaped noise. Preliminary results show that normal-hearing listeners benefited from Col-SE processing in conjunction with minimal phonemically targeted speech processing more than with either processing strategy individually. Increases in consonant amplitude and duration beyond a modest amount reduce intelligibility. These data suggest that the two processing strategies are viable to improve speech intelligibility, but that there is a limit to the processing whereby benefits are no longer observed.

2aPP15. Effects of independent bilateral compressive amplification on lateralization of a single source. Ian M. Wiggins and Bernhard U. Seeber (MRC Inst. of Hearing Res., Nottingham NG7 2RD, United Kingdom, ian@ihr.mrc.ac.uk)

Use of compressive amplification in bilateral hearing aid fittings can disrupt binaural cues important to spatial hearing. The head-related transfer function introduces direction-dependent interaural time and level differences (ITDs and ILDs) which are consistent with one another. Independent bilateral compression, however, reduces ILDs such that they suggest a different, conflicting direction than ITDs. The reduction in ILDs is a dynamic effect that depends on the characteristics of the compression and the sound. Single-channel compression was applied over a wide, high-frequency band. Two conditions were run, with the high-frequency channel presented to listeners in isolation or recombined with an unprocessed low-frequency channel. Stimuli included pink noise bursts with varied onset slopes and rates of amplitude modulation and speech. The effects of compression on the perceived auditory objects were assessed using a semantic differential method, in which listeners rated various spatial attributes on scales between two bipolar adjectives, for example, image width was rated on a scale between "focused" and "diffuse." Additionally, a lateralization task was performed. Initial results show that for sounds with slow onsets, compression tends to shift image location or causes the image to split. Interestingly, image location can also be affected for speech, particularly if high-pass filtered.

2aPP16. Horizontal localization and hearing in noise ability in adults with sensorineural hearing loss using hearing aids with binaural processing. Amy R. Mullin (Dept. of Commun. Sci. and Disord., The Univ. of Texas at Austin, 1 University Station A1100, Austin, TX 78712, amyruthie@mail.utexas.edu)

The purpose of the study was to determine whether hearing aids with binaural processing improve performance during a localization and a hearing in noise task. The study included 15 participants between the ages of 29 and 68 who had a bilateral symmetrical sensorineural hearing loss and who had no prior hearing aid experience. Participants were fitted with Oticon Epoq XW receiver-in-the-ear hearing aids bilaterally. The participants completed a horizontal localization task and a hearing in noise task with three listening conditions: (1) without hearing aids (NO), (2) with hearing aids that were not linked (BIL), and (3) with hearing aids that were linked (BIN). For the horizontal localization task, 1.5-s pink noise bursts were used as the stimulus. Sentences from the Hearing in Noise Test were used as target stimuli for the hearing in noise task. Continuous discourse by one male and two female talkers was recorded and used as maskers. The specific aim of the localization and hearing in noise tasks was to determine which of the listening conditions resulted in the best score for each task. Data are still being collected and data analysis will follow.

2aPP17. Effect of waveform shape and polarity on loudness and on place pitch for cochlear implant users. Robert Carlyon, Olivier Macherey, and John Deeks (MRC Cognition and Brain Sci. Unit, 15 Chaucer Rd., Cambridge CB2 7EF, United Kingdom)

It has been shown that, for pseudomonophasic pulses presented in monopolar mode, less current is needed to achieve the same loudness when the short-high phase is anodic than when it is cathodic. Experiment 1 extended that finding, obtained with the Advanced Bionics (AB) device, to waveforms that can be implemented in other devices, using 99-pps, 32-μs/phase pulse trains. For AB and MedEl devices, stimuli were triphasic pulses whose central phase had twice the amplitude of the first and third phases. For the Nucleus device, stimuli were pairs of biphasic pulses with opposite leading polarity, each with a 58-μs interphase-gap, separated by 8 μs—resulting in two adjacent same-polarity phases in the center of the waveform. The current level needed was always 1–2 dB lower when the central portion was anodic than when it was cathodic—including for electrode 1 of the MedEl device, which is inserted deep into the cochlea. Experiment 2 showed that, for pseudomonophasic pulses in bipolar mode, (a) pitch is lower when the "short-high" pulse is anodic relative to the more-apical than to the more-basal electrode and (b) intermediate pitches can be produced by variations in waveform shape and polarity applied to the same bipolar pair.

2aPP18. Pitch-ranking of electric and acoustic stimuli by cochlear implant users with the HiRes and Fidelity120 speech processing strategies. Benjamin A. Russell and Gail S. Donaldson (Dept. of Comm. Sci. and Disord., Univ. of South Florida, Tampa, FL 33620, barussel@mail.usf.edu)

Estimates of place-pitch sensitivity in cochlear implant (CI) users are typically obtained using electric pulse trains presented directly to the implanted electrodes. Such estimates may overestimate the place-pitch sensitivity available to these listeners through their speech processors due to spectral smearing by the analysis filters. To determine the influence of speech processing on place-pitch sensitivity, electric pitch-ranking (EPR) and acoustic pitch-ranking (APR) thresholds were compared in four users of the Advanced Bionics CI. EPR thresholds were obtained for single- and dual-electrode pulse-train stimuli presented to electrodes near the center of the array. APR thresholds were obtained for pure tones having frequencies corresponding to the tonotopic locations of the electric pulse trains, using both the HiRes and Fidelity120 speech processing strategies. Counter to expectation, APR thresholds were similar to EPR thresholds for three of four subjects, and APR thresholds were smaller than EPR thresholds for the remaining subject. APR thresholds did not differ systematically for the HiRes and Fidelity120 strategies, consistent with Nogueira *et al.* [(in press). EURASIP J Adv Signal Proc]. Findings suggest that CI users can make use of across-channel cues when performing an APR task and that such cues can compensate for spectral smearing by the analysis filters.

2aPP19. Lowering mean fundamental frequency to improve speech intelligibility in noise under simulated electric-acoustic stimulation. Christopher A. Brown and Sid P. Bacon (Dept. of Speech and Hearing Sci., P.O. Box 870102, Tempe, AZ 85287-0102)

We have previously demonstrated that much or all of the benefits of electric-acoustic stimulation (EAS) can be achieved when the low-frequency speech is replaced with a tone that is modulated in both frequency and amplitude with F0 and amplitude envelope cues derived from the target speech. One advantage of this approach is that the frequency of the carrier tone can

be lowered with little decline in benefit. Lowering mean F0 has the potential to provide EAS benefit to CI users who have very limited residual hearing. One drawback to this approach is that it relies on the efficacy of the pitch extraction algorithm. This is problematic because pitch extractors have trouble in background noise, an environment in which F0 is particularly useful. Here, an alternative way of lowering mean F0 that is unaffected by the presence of noise is examined under simulated EAS conditions. Speech intelligibility was measured using an algorithm based on resampling and compared to performance with the pitch-based method we have used previously, at frequency shifts of 0, 0.5, and 1 octave. At the 0.5-octave shift, the resampling-based approach provided more benefit than the tone. However, at the 1-octave shift, resampling was less beneficial. [Work supported by NIDCD.]

2aPP20. Effect of auditory deprivation on binaural sensitivity in bilateral cochlear implant users. Ruth Y. Litovsky, Gary L. Jones, and Richard VanHoesel (Waisman Ctr., Univ. of Wisconsin, 1500 Highland Ave., Madison, WI 53705, litovsky@waisman.wisc.edu)

Increasing numbers of cochlear implant (CI) users receive Implants in both ears in an effort to restore spatial hearing abilities. A known limitation of commercial CI devices is lack of synchrony to the signal fine timing cues arriving at the two ears. In addition to these hardware limitations, which may contribute in substantial ways to the gap in performance typically seen between bilateral CI users and normal-hearing people, there are considerable inter-subject differences in the type and history of hearing loss. We use a research processor which enables inter-aurally coordinated pulsatile stimulation of selected pairs of electrodes in the right and left ears. This project is concerned with the effects of the age at onset of deafness, and place of stimulation, on binaural sensitivity. Thresholds for discrimination of inter-aural time difference (ITD) and inter-aural level difference(ILD) and pointer-identification for perceived intracranial position were measured. Prelingually deafened subjects had no sensitivity to ITD but retained ILD sensitivity. People with childhood- or adult-onset of deafness retained sensitivity to both ITD and ILD, some within normal-hearing level of performance on some conditions. The role of auditory deprivation in the emergence and preservation of binaural sensitivity will be discussed. [Work supported by NIH-NIDCD.]

TUESDAY MORNING, 20 APRIL 2010 HARBORSIDE B, 8:20 TO 10:15 A.M.

Session 2aSAa

Structural Acoustics and Vibration, Noise, Underwater Acoustics, Animal Bioacoustics, and INCE: Noise Control of Small Marine Vehicles

Joseph M. Cuschieri, Chair
Lockheed Martin Corp., 100 East 17th St., Riviera Beach, FL 33404

Chair's Introduction—8:20

Invited Papers

8:25

2aSAa1. An overview of unmanned underwater vehicle noise in the low to mid frequency bands. Jason D. Holmes (Raytheon BBN Technologies, 10 Moulton St., Cambridge, MA 02138, jholmes@bbn.com), William M. Carey (Boston Univ., Boston, MA 02215), and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Unmanned (autonomous) underwater vehicles offer a unique, cost-effective platform for performing ocean acoustic measurements and surveys because multiple systems can be deployed from a single research vessel. Various data surveys can be performed including on-the-bottom geo-acoustic surveys over large areas, sub-sea-surface turbulence and microbubble structure surveys, and bi-static fish population surveys. To take advantage of the autonomous survey capabilities of underwater vehicles, sufficient signal-to-noise ratio and acoustic aperture (resolution) are required for acoustic measurements. The most commonly used vehicle sonar systems provide images utilizing high frequency hull mounted arrays and sources. In the lower-frequency band (100 Hz–10 kHz), however, vehicle noise levels and aperture remain the two most significant challenges, especially for passive systems. Previous experimental and analytical work has shown that a towed array with synthetic aperture processing can be used to obtain the necessary aperture. The major challenge of vehicle radiated noise is the focus of this paper, and both measured and archival results on vehicle noise are presented including an overview of levels, spectral character, and noise mechanisms for several vehicles. In particular, fundamental vehicle propulsion system noise is discussed along with implications on measurement performance and possible mitigation strategies.

8:45

2aSAa2. Unmanned underwater vehicle self-noise and implications for low-frequency sonar design. Brian Houston (Naval Res. Lab., Code 7130, 4555 Overlook Ave., Washington, DC 20375)

Sonars configured on small unmanned underwater vehicles tend to be higher-frequency systems that are relatively insensitive to the dominating self-noise associated with low-frequency energy sources. These include the main propulsor, control surface actuators, and navigation components. A new generation of vehicles is required having reduced low-frequency signatures in order to support the growing interest in low-frequency broadband active and passive sonars. In the work presented here, we discuss the impact of vehicle noise

on the performance of both real and synthetic receiver arrays. Self-noise mitigation methods and their impact on vehicle architecture will be discussed as well as the use of signal processing techniques employing acoustic holographic projection to reduce the impact of structural-borne noise. [Work supported by ONR.]

9:05

2aSAa3. Decreasing the radiated acoustic and vibration noise of a mid-size, prop-driven, autonomous underwater vehicle. Richard Zimmerman, Gerald D'Spain, and John Orcutt (Scripps Inst. of Oceanogr., Univ. California San Diego, 291 Rosecrans St., San Diego, CA 92106, rzimmerman@ucsd.edu)

Previously published efforts at decreasing the radiated acoustic and vibration noise of the propulsion system of an Odyssey IIb autonomous underwater vehicle (AUV) manufactured by Bluefin Robotics, Inc. resulted in noise levels recorded by an AUV-mounted hydrophone array that were at or below typical background ocean noise levels across much of the frequency band from 300 Hz to 10 kHz [IEEE J. Ocean. Eng. **30**, 179–187]. The modifications required to achieve this 20–50-dB reduction in propulsion noise levels will be reviewed in this talk. Recently, these modifications have been incorporated into the Bluefin 21 AUV at the Scripps Institution of Oceanography. In addition, the stepper motors in the linear actuators used to steer this AUV's vectored-thrust tail cone in depth and heading have been replaced. At-sea measurements show that the high-level, broadband transients that previously occurred every 2–3 s due to these actuators are no longer visible in the hydrophone data. Eliminating these sources of self-noise allow the vehicle to be used in marine biological studies without vehicle noise disturbing either the acoustic measurements themselves or the habitats under observation. [Work supported by the Office of Naval Research and British Petroleum.]

9:25

2aSAa4. Acoustic noise estimates for a quiet unmanned underwater vehicle. Carl A. Cascio (50 Myrock Ave., Waterford, CT 06385-3008)

Acoustical Technologies Inc. (ATI) has been tasked by a number of clients to predict the radiated noise of unmanned underwater vehicles (UUVs). Laboratory, dockside, and at sea measurements have been taken on 21-in.-diameter, 20-ft-long prototype vehicles. Data were obtained during operation of the propulsion motors, control surfaces, and other UUV components. Structural impact hammer testing was also used to estimate propagation path transfer functions. Acoustic noise and structural vibration sensors were recorded on the ATI multi-channel data acquisition system and analyzed to produce narrow and broad bandwidth spectral estimates over the 10-Hz–100-KHz frequency range. Radiated noise predictions from the component source level data and transfer functions were made and illustrate the importance of selecting quiet components and appropriate noise control. This paper takes a look at how quiet a 21-in.-diameter 20-ft-long UUV could be using what is currently known about typical UUV component hardware. This is not an estimate for a real UUV but a radiated noise model based on an ATI concept using quiet components, practical noise control, and generic transfer functions. Radiated noise estimates will be presented at several operating speeds.

Contributed Papers

9:45

2aSAa5. Turbulent boundary layers over hydrophone arrays. Craig N. Dolder (Dept. of Mech. Eng., The Univ.of Texas at Austin, 1 University Station C2200, Austin, TX 78712-0292, dolder@mail.utexas.edu), Michael R. Haberman, and Charles E. Tinney (The Univ. of Texas at Austin, Austin, TX 78712-0235)

The speed at which naval vessels can operate sonar receiver arrays is limited due to the noise produced by the formation of a turbulent boundary layer (TBL) over the hull of the vessel. Despite efforts in the signal processing community to reduce these signatures, flow noise levels continue to surpass the signal to noise ratio needed for effective sonar operation at high-vessel speeds. This study focuses on the hydrodynamic signatures induced by the turbulence and the means by which energy can be removed from the TBL structures to reduce pressure fluctuations on the array elements. The research presented here employs an array of hydrophones with high-spatial resolution to measure the dynamic surface pressure while simultaneously acquiring time resolved single-point velocity field measurements at various positions above the array and within the TBL using laser Doppler velocimetry. Classical statistical quantities are then computed to determine the relationship between the pressure and velocity fields. Further, an analysis of the wave number frequency makeup of the surface pressure provides insight into the convective nature of the pressure filtered flow structures. The findings provide helpful guidance in developing active control methods for reducing the TBL noise.

10:00

2aSAa6. Estimation and measurement of the acoustic signature of unmanned surface and underwater vehicles. Joseph Cuschieri (Lockheed Martin MS2, Undersea Systems, Riviera Beach, FL 33404)

As the use of unmanned surface and underwater vehicles (USVs and UUVs) increases, the acoustic signature and self-noise of these vehicles become important for certain type of applications. For low-detection probability the UUV or USV must have a low-acoustic signature. Furthermore, the self-noise of the UUV or USV may interfere with on-board sonar sensors, especially as more sophisticated sonar systems are being developed. As the concept of the USV or UUV is to design for low cost and low weight, with low availability of power, it is important to address the mitigation of the acoustic signature during the design phase through estimation techniques based on modeling and component level selection. Additionally during the test and integration phase, methods to measure the acoustic signature without resorting to expensive offshore measurements in "quiet" acoustic test ranges are required. In this presentation some of the approaches available for modeling and testing are discussed, together with the challenges faced by noise control engineers working these type problems and the type of noise sources that have to be addressed.

Session 2aSAb

Structural Acoustics and Vibration and INCE: Space Vehicle Vibroacoustics

Dean E. Capone, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Stephen C. Conlon, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Invited Papers

10:30

2aSAb1. Vibroacoustics in airplane design. Evan B. Davis (The Boeing Co., P.O. Box 3707, MC 67-ML Seattle, WA 98124, evan.b.davis@boeing.com)

Structural and structural-acoustic (vibroacoustic) tools are used to predict the future performance of systems in order to guide design trades and system optimization decisions. The key questions to be answered are (1) what needs to be known? (2) how well does it need to be known? and (3) how reliable are the tools that will be used to make the decisions?

10:50

2aSAb2. Vibro-acoustic analysis of aerospace structures and issues with the available commercial prediction tools. Ali R. Kolaini and Dennis L. Kern (Jet Propulsion Lab., California Inst. of Technol., 4800 Oak Grove Dr., Pasadena, CA 91109-8099)

The results of vibro-acoustic modeling using the boundary element method (BEM) that predicts the acceleration responses at critical locations and at the interfaces of selected test articles are discussed. High fidelity acoustic tests were performed in a couple of cases and the results are used to validate the BEM predictions. The accuracy of the BEM and its ability to correctly predict the acceleration responses of lightweight structures are discussed in some detail. Also a combined system level BEM, consisting of structures very responsive to acoustic pressures, and force-limited base shake random vibration analysis was performed. We will discuss how these results may be used to derive random vibration specifications for the purpose of qualifying large and lightweight structures for flight. In this paper, we also discuss the commercially available vibro-acoustic tools that are used to predict the acoustic transmission losses and vibration responses of flight structures for lift-off (assumed to be diffuse) and transonic (turbulent boundary layer) acoustic fields. The pros and cons of using the statistical energy analysis, finite element analysis, BEM, and newly developed hybrid methods within these vibro-acoustic tools are discussed in some detail.

Contributed Paper

11:10

2aSAb3. Noise control in Space Shuttle Orbiter. Jerry R. Goodman (SF/Habitability & Enviromental Factors Div., NASA-JSC, Houston, TX 77058)

Acoustic limits in habitable space enclosures are required to ensure crew safety, comfort, and habitability. Noise control is implemented to ensure compliance with the acoustic requirements. The purpose of this paper is to describe problems with establishing acoustic requirements and noise control efforts, and to present examples of noise control treatments and design applications used in the Space Shuttle Orbiter. Included is the need to implement the design discipline of acoustics early in the design process and noise control throughout a program to ensure that limits are met. The use of dedicated personnel to provide expertise and oversight of acoustic requirements and noise control implementation has shown to be of value in the Space Shuttle Orbiter program. It is concluded that to achieve acceptable and safe noise levels in the crew habitable space, early resolution of acoustic requirements and implementation of effective noise control efforts are needed. Management support of established acoustic requirements and noise control efforts is essential.

Session 2aSC

Speech Communication: Speech and Noise

Carol Y. Espy-Wilson, Chair
Univ. of Maryland, Electrical and Computer Engineering, A. V. Williams Bldg., College Park, MD 20742

Chair's Introduction—8:00

Invited Papers

8:05

2aSC1. Noise-suppression algorithms for improved speech intelligibility by normal-hearing and cochlear implant listeners. Philipos Loizou (Dept. of Elec. Eng., Univ. of Texas-Dallas, Richardson, TX 75080, loizou@utdallas.edu)

Much research in the past few decades focused on the development of noise reduction algorithms that can suppress background noise. While these single-microphone based algorithms have been proven to improve the subjective speech quality, they have not been effective in improving speech intelligibility. This is partly due to the fact that most noise-suppression algorithms introduce speech distortion and partly because most algorithms are not optimized to operate in a particular noisy environment. Furthermore, none of the existing noise-reduction algorithms was designed to optimize a metric that correlates highly with intelligibility. This talk will present intelligibility data collected with normal-hearing and cochlear implant listeners who were presented with noisy speech processed by environment-optimized algorithms. It will also present algorithms that were designed using metrics that correlate highly with speech intelligibility. The data from these studies suggest that it is possible to develop noise reduction algorithms that improve speech intelligibility provided some constraints are imposed on the design of the suppression function and/or the intended listening environment. Research supported by NIDCD/NIH.]

8:25

2aSC2. Subjective evaluation of the speech quality from speech enhancement and segregation algorithms. Vijay Mahadevan, Srikanth Vishnubhotla, and Carol Espy-Wilson (Univ. of Maryland, 3180, A V W Bldg., College Park, MD 20770)

Automatic separation of speech from noise and segregation of overlapping co-channel speech are two of the most challenging problems in speech processing. In previous work, we have developed algorithms to both enhance noisy speech and segregate overlapping speech streams. Our single-channel speech enhancement and speech segregation algorithms have shown better performance than other reported algorithms for automatic speech recognition. Additionally, objective evaluation scores of perceptual quality have shown a significant improvement following processing by our algorithms. In this study, we focus on subjective evaluation of these algorithms for human listeners. We investigate the intelligibility and quality of speech from our algorithm on normal-hearing listeners, cochlear implant users, and hearing impaired subjects. Our preliminary results indicate a significant improvement in the perceptual quality of the speech signal after being processed by our algorithm and suggest that the proposed algorithms can be used as a pre-processing block within the signal processing in hearing aid devices.

8:45

2aSC3. Speech recognition in loud noise using an ear-insert microphone. Tarun Pruthi, Mihai Despa, and Amit Juneja (Think A Move, Ltd., 23715 Mercantile Rd. Ste. 100, Beachwood, OH 44122)

Think A Move, Ltd. has developed a patented ear-insert microphone which captures speech as acoustic vibrations inside the ear canal. These vibrations propagate to the ear canal through the flesh and bones in the human skull. A high density foam on the earpiece seals the ear canal when the earpiece is inserted. Tests show that this earpiece provides an average passive noise cancellation (PNC) of around 38 dB for noises in Aurora database. Using an in-house speech command recognizer, with a short enrollment phase, on a database of 19 speakers (11 females, 8 males), speaking a vocabulary of 56 commands, an average accuracy of 85% has been observed in 90 dBA of tank, military vehicle, and machine gun noises. To further demonstrate the noise robustness of the earpiece as compared to external microphones, pilot tests were conducted on a small set of speakers to recognize speech commands recorded simultaneously with an external microphone and our ear-insert microphone with our recognizer. Results show that while the accuracy of the recognizer drops to 27% in 90 dBA noise from 96% in quiet for external microphone, it only drops to 92% in 90 dBA of noise from 95% in quiet for internal microphone.

9:05

2aSC4. Voice conversion for enhancing various types of body-conducted speech detected with non-audible murmur microphone. Tomoki Toda (Graduate School of Information Sci., Nara Inst. of Sci. and Technol., Takayama-cho 8916-5, Ikoma-shi, Nara 630-0192 Japan, tomoki@is.naist.jp)

Our proposed statistical voice conversion approach to enhancing various types of body-conducted speech detected with Non-Audible Murmur (NAM) microphone is presented in this talk. NAM microphone, one of the body-conductive microphones [Nakajima *et al.*, IEICE Trans. Inf. and Syst., **E89-D**, 1–8 (2006)], enables us to detect various types of body-conducted speech such as extremely

soft whisper, normal speech, and so on. Moreover, it is robust against external noise due to its noise-proof structure. To make speech communication more universal by effectively using these properties of NAM microphone, body-conducted speech enhancement techniques have been developed with a state-of-the-art statistical voice conversion algorithm [Toda *et al.*, IEEE Trans. ASLP, **15**, 2222–2235 (2007)]. The proposed techniques would bring a new paradigm to human-to-human speech communication: e.g., the use of body-conducted voiced speech for noise robust speech communication, the use of body-conducted unvoiced speech for silent speech communication, and the use of body-conducted artificial speech for speaking aid [T Toda *et al.*, Proc. ICASSP (2009) pp. 3601–3604]. This talk gives an overview of these promising techniques and presents their applications. [This research was supported in part by MIC SCOPE.]

9:25

2aSC5. Synthesizing speech from surface electromyography and acoustic Doppler sonar. Arthur R. Toth (Yap Inc., 2414 Shady Ave., Pittsburgh, PA 15217, atoth@cs.cmu.edu), Michael Wand (Univ. Karlsruhe, Karlsruhe, Baden-Württemberg 76131, Germany), Szu-Chen Stan Jou (ATC, ICL, Industrial Technol. Res. Inst., Chutung, Hsinchu, Taiwan 31040), Tanja Schultz (Univ. Karlsruhe, Karlsruhe, Baden-Württemberg 76131, Germany), Bhiksha Raj (Carnegie Mellon Univ., Pittsburgh, PA 15213), Kaustubh Kalgaonkar (Georgia Inst. of Technol., Atlanta, GA 30332), and Tony Ezzat (Mitsubishi Electric Res. Labs., Cambridge, MA 02139)

Numerous techniques have been devised to process speech audio in noise, but automatic speech recognition is difficult when the noise is too great. An alternative approach is to collect data that represent the speech production process but is less affected by noise in the speech audio range. Two such types of data come from surface electromyography (EMG) and acoustic Doppler sonar (ADS). EMG records muscle activation potentials. ADS records reflected ultrasound tones. Both can be used to measure facial movements related to speech, but they present their own challenges for automatic speech recognition. This work investigates the alternative approach of using these data sources for speech synthesis. The synthesis techniques explored in this work are based on Gaussian mixture model mapping techniques, which are commonly used for voice transformation. Voice transformation is traditionally concerned with changing the identity of speech audio signals, but others have demonstrated that such techniques can be used to transform different types of signals, such as non-audible murmur and electromagnetic articulography, to speech. This work demonstrates that such techniques also show promise for transforming EMG and ADS signals to speech.

9:45

2aSC6. Dealing with noise in automatic speech recognition. Douglas O'Shaughnessy (INRS-EMT, Univ. of Quebec, 800 de la Gauchetiere West, Ste. 6900, Montreal QC H5A 1K6, Canada)

While automatic speech recognition (ASR) can work very well for clean speech, recognition accuracy often degrades significantly when the speech signal is subject to corruption, as occurs in many communication channels. This paper will survey recent methods for handling various distortions in practical ASR. The problem is often presented as an issue of mismatch between the models that are created during prior training phases and unforeseen environmental acoustic conditions that occur during the normal test phase. As one can never anticipate all possible future conditions, ASR analysis must be able to adapt to a wide variety of distortions. Human listeners furnish a useful standard of comparison for ASR in that humans are much more flexible in handling unexpected acoustic distortions than current ASR is. Methods that adapt ASR features and models will be compared against ASR methods that enhance the noisy input speech. Other topics to be discussed will include estimation of noise and channel parameters, RASTA, and cepstral mean normalization. TRAP-TANDEM features Vector Taylor Series, joint speech and noise modeling, and advanced front-end feature extraction. Single-microphone versus multi-microphone approaches will also be discussed.

10:05—10:25 Break

10:25

2aSC7. Using speech models for separation in monaural and binaural contexts. Daniel P. Ellis, Ron J. Weiss, and Michael I. Mandel (Dept. of Elec. Eng., Columbia Univ., 500 W. 120th St., Rm. 1300, New York, NY 10027, dpwe@ee.columbia.edu)

When the number of sources exceeds the number of microphones, acoustic source separation is an underconstrained problem that must rely on additional constraints for solution. In a single-channel environment the expected behavior of the source—i.e., an acoustic model—is the only feasible basis for separation. We have developed an approach to monaural speech separation based on fitting parametric "eigenvoice" speaker-adapted models to both voices in a mixture. In a binaural, reverberant environment, the interaural characteristics of an acoustic source exhibit structure that can be used to separate even without prior knowledge of location or room characteristics. For this scenario, we have developed MESSL, an EM-based system for source separation and localization. MESSL's probabilistic foundation facilitates the incorporation of more specific source models; MESSL-EV incorporates the eigenvoice speech models for improved binaural separation in reverberant environments.

10:45

2aSC8. Dual stage probabilistic voice activity detector. Ivan Tashev (Microsoft Res., One Microsoft Way, Redmond, WA 98052), Andrew Lovitt, and Alex Acero (Microsoft Res., Redmond, WA 98052)

Voice activity detectors (VADs) are critical part of every speech enhancement and speech processing system. One of the major problems in practical realizations is to achieve robust VAD in conditions of background noise. Most of the statistical model-based approaches employ the Gaussian assumption in the discrete Fourier transform domain, which deviates from the real observation. In this paper, we propose a class of VAD algorithms based on several statistical models of the probability density functions of the magnitudes. In addition, we evaluate several approaches for time smoothing the magnitude response to achieve a more robust estimate. A large data corpus of in-car noise conditions is then used to optimize the parameters of the VAD, and the results are discussed.

11:05

2aSC9. Speech enhancement beyond minimum mean squared error with perceptual noise shaping. Lae-Hoon Kim, Kyung-Tae Kim, and Mark Hasegawa-Johnson (Dept. of ECE, Univ. of Illinois at Urbana-Champaign, 405 North Mathews Ave., Urbana, IL 61820, jhasegaw@illinois.edu)

Residual error signal after speech enhancement through linear filtering can be decomposed into two disjoint portions: speech signal distortion and background noise suppression. Speech is known to follow a super-Gaussian probabilistic distribution function (PDF) such as Laplacian, while background noise follows Gaussian PDF. Minimum mean squared error estimation requires only second order statistics not only for the noise but also for the speech. Therefore higher-order dependence of observed speech on the original speech may cause leakage of speech information into the error residual. This talk will formulate an optimization problem minimizing higher-order statistics (HOS) as well as energy of the signal distortion constrained by a limit on the maximum audibility of the residual noise. Note that due to the non-stationary nature of speech, we perform the speech enhancement in short overlapping frames. Minimizing HOS of the speech distortion ensures that the speech distortion includes only noise terms, with minimum leakage from the speech signal. The constraint on the residual noise margin prevents over-suppressing, which may result in unwanted speech distortion.

11:25

2aSC10. The role of temporal modulation processing in speech/non-speech discrimination tasks. Hong You and Abeer Alwan (Dept. of Elec. Eng., UCLA, 405 Hilgard Ave., Los Angeles, CA 90095, alwan@ee.ucla.edu)

In this paper, temporal modulation characteristics of speech and noise from the point of view of speech/non-speech discrimination are analyzed. Although previous psychoacoustic studies have shown that temporal modulation components below 16 Hz are important for speech intelligibility, there is no reported analysis of modulation components from the point of view of speech/noise discrimination. Our data-driven analysis of modulation components of speech and noise reveals that speech and noise are more accurately classified by low-pass modulation frequencies than band-pass ones [H. You and A. Alwan, in Interspeech Proceedings (2009) pp. 36–39]. Effects of additive noise on the modulation characteristics of speech signals are also analyzed. Based on the analysis, a frequency adaptive modulation processing algorithm for a noise robust automatic speech recognition task is proposed. Speech recognition experiments are performed to compare the proposed algorithm with other noise robust front-ends, including RASTA and ETSI-AFE. Recognition results show that the frequency adaptive modulation processing algorithm is promising and is of low complexity. [Work supported in part by NSF.]

Contributed Paper

11:45

2aSC11. Relevant spectro-temporal modulations for robust speech and nonspeech classification. Sridhar Krishna Nemala and Mounya Elhilali (Dept. of Elec. & Comput. Eng., 3400 N. Charles St., Barton 105, Baltimore, MD 21218)

Robust speech/non-speech classification is an important step in a variety of speech processing applications. For example, in speech and speaker recognition systems designed to work in real world environments, a robust discrimination of speech from other sounds is an essential pre-processing step. Auditory-based features at multiple-scales of time and spectral resolution have been shown to be very useful for the speech/non-speech classification task [Mesgarani *et al.*, IEEE Trans. Speech Audio Process. **10**, 504–516 (2002)]. The features used are computed using a biologically inspired auditory model that maps a given sound to a high-dimensional representation of its spectro-temporal modulations (mimicking the various stages taking place along the auditory pathway from the periphery all the way to the primary auditory cortex). In this work, we analyze the contribution of different temporal and spectral modulations for robust speech/non-speech classification. The results suggest the temporal modulations in the range 12–22 Hz, and spectral modulations in the range 1.5–4 cycles/octave are particularly useful to achieve the robustness in highly noisy and reverberant environments.

TUESDAY MORNING, 20 APRIL 2010 GALENA, 9:10 TO 11:30 A.M.

Session 2aSP

Signal Processing in Acoustics and Underwater Acoustics: Arrays in Water and Air

Charles F. Gaumond, Chair
Naval Research Lab., 4555 Overlook Ave., Code 7142, Washington, DC 20375

Chair's Introduction—9:10

Contributed Papers

9:15

2aSP1. Estimating the size and spatial distribution of bubble clouds in an underwater acoustic test tank. Fred D. Holt, IV (Appl. Res. Lab. and Graduate Program in Acoust., The Penn State Univ., State College, PA 16804), J. Daniel Park, R. Lee Culver (The Penn State Univ., State College, PA 16804), David Coles, and Timothy Leighton (Univ. of Southampton, Southampton S017 1BJ, United Kingdom)

The AB Wood Underwater Acoustic test tank at ISVR, University of Southampton, UK is outfitted with a Venturi-based apparatus for generating bubble clouds containing a large range of bubble sizes. Larger bubbles are allowed to rise out in a settling tank before the nearly opaque bubbly water is pumped into the acoustic test tank through a discharge manifold on the bottom. A series of acoustic attenuation measurements was made at AB Wood in July 2008 in order to estimate the size and spatial distribution of bubbles in the cloud. An acoustic projector transmitted 1-ms pure tones from 25 to 100 kHz, which were received at a co-linear three-element horizontal array. The receive hydrophones were spaced 0.5 m apart, with the center hydrophone placed directly over the bubble discharge location. Measurements were taken on-axis with the projector, and at three distances off-axis,

in 12-in. increments. The attenuation measurements were used to estimate the size and spatial distribution of bubbles within the cloud. The bubble size data will be used to support a study of the effect of nearby bubbles on array performance degradation. [Work sponsored by the Office of Naval Research, Code 321.]

9:30

2aSP2. Measurement and analysis of array gain degradation due to bubble scattering. J. Daniel Park, Fred D. Holt, IV, R. Lee Culver (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804), David Coles, and Timothy Leighton (Univ. of Southampton, Southampton, United Kingdom)

When an array is steered in the direction of a signal, array gain (AG) is maximum when the signal is coherent across the array, meaning that the signals add in phase for all elements and the noise or interference is incoherent across the array (i.e., it adds with random phase). For an acoustic array operating in the ocean, we would like to understand the degree to which scattering by nearby bubbles degrades AG. Bubble attenuation can also degrade array performance by attenuating the signal of interest, but that is separate from AG degradation. The degradation of AG due to scattering by nearby bubbles has been measured for different bubble densities. We have analyzed the relationship between bubble density at the array and the degradation in AG. We present probability density functions (pdfs) of signal amplitude and phase at the array elements. The amplitude pdfs can be approximated as a Rayleigh–Rice distribution, and the phase pdfs follow the von Mises distribution. With independence assumed between amplitude and phase distributions, simulated distributions generate array gain distributions closely matching the measured AG distribution. [Work sponsored by ONR Undersea Signal Processing.]

9:45

2aSP3. Array performance in a complex littoral environment. Steven L. Means (Naval Res. Lab., Code 7120, 4555 Overlook Ave. SW, Washington, DC 20375, steve.means@nrl.navy.mil), Richard M. Heitmeyer (Global Strategies Group Inc., Washington, DC 20375), and Stephen C. Wales (Naval Res. Lab., Washington, DC 20375)

In August of 2007 a long (~1-km), 500 phone linear array began collecting acoustic data in the waters (~260 m in depth) approximately 12 km off the coast of Fort Lauderdale, FL. The array consisted of four, 125-phone segments deployed closely along a line running nearly east and west. Marine-band radar data were collected concurrently so that shipping in the region of the array could be tracked. The data considered in this analysis were obtained over the month of August and contains ~19 days of measurements. Array performance is investigated by beamforming at a number of frequencies (up to ~420 Hz) and apertures then determining cumulative distribution functions as a function of bearing and noise window statistics. The results are compared for day, night, weekday, and weekend measurements during which the local shipping varies significantly. Additionally, ship tracks obtained from the acoustic array are compared against those obtained from radar. [Work supported by ONR base funding at NRL.]

10:00

2aSP4. Results of hermetic transform signal processing to enhance the resolution and array gain of underwater acoustic arrays. Harvey C. Woodsum (Bayshore Labs Div., Sonetech Corp., 10 Commerce Park North Unit 1, Bedford, NH 03110)

Results of applying the discrete hermetic transform (DHT) to the beamforming of underwater acoustic arrays are presented in terms of measured reductions in beam mainlobe width as well as the associated improvement in directivity index/array gain for practical cases of interest. Application of the DHT to array beamforming is shown to produce substantially enhanced resolution relative to the conventional beamforming diffraction limit as well as significant enhancement in array gain against ambient noise, without the use of data adaptive or nonlinear processing. As a result, enhanced sonar signal detection and/or the ability to use substantially smaller than normal arrays can be accomplished through the judicious use of DHT based beamforming algorithms. Results are favorably compared to theoretical predictions of algorithm performance.

10:15—10:30 Break

10:30

2aSP5. A mobile acoustic multiple-input/multiple-ouput communication testbed. Aijun Song, Mohsen Badiey, and Arthur Trembanis (College of Earth, Ocean, and Environment, Robinson Hall-114, Newark, DE 19716)

Underwater acoustic data communication is critical for naval and scientific underwater missions. For example, high-rate telemetry between underway autonomous underwater vehicles (AUVs) and surface platforms can facilitate adaptive sampling of the ocean. Multiple-input/multiple-output (MIMO) systems can deliver significant increased channel capacity for underwater communications. At the University of Delaware, an acoustic MIMO communication testbed on a Gavia AUV has been developed for digital communication measurements at the frequency band of 20–30 kHz. The Gavia AUV is a modular, small-size vehicle (20 cm in diameter, 77 kg in air) with a depth rating of 200 m. With advanced navigation systems (INS) and surface communication capabilities (WiFi and Iridium), it has been tested through various coastal-ocean missions. The MIMO system on board is designed for easy, low-cost deployment, and is capable of conducting acoustical sampling via a towed array with or without single-element or multi-element source transmission. In the presentation, the design concept, component detail, and engineering test results will be shown. [Work supported by ONR 321OA.]

10:45

2aSP6. Experimental validation of Helmholtz equation least squares when the Nyquist spatial sampling requirement is violated. Richard E. Dziklinski, III and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Previous numerical studies [Dziklinsky and Wu, J. Acoust. Soc. Am. (2009)] have shown that Helmholtz equation least squares (HELSs) enable one to violate the spatial Nyquist sampling frequency and stand-off distance requirements inherent in Fourier acoustics. A direct benefit of this severe under sampling using HELS is a significant saving in measurement and computational effort in locating acoustic point sources in practice without loss of the required spatial resolution. The presented paper aims at validating the previous numerical studies by conducting experiments on locating two incoherent point sources separated by a small distance of 6.25 mm via HELS. The acoustic pressures are measured by a 5×5 microphone array over a 50×50 mm square plane at varying stand-off distances with fixed microphone spacing of 12.5 mm. The considered source frequencies are well above the spatial Nyquist sampling frequency. Results show that HELS is capable of locating point sources when stand-off distance is 10 mm or less and SNR is 10 dB or higher. Similar results obtained by using planar Fourier acoustics are also presented for comparison purposes.

11:00

2aSP7. Implementation of sound ball with acoustic contrast control method. Min-Ho Song (Grad. School of Culture Technol., KAIST, Sci. Town, Daejon 305-701, Korea, godspd@kaist.ac.kr) and Yang-Hann Kim (KAIST, Sci. Town, Daejon 305-701, Korea)

It is well known that the problem of generating sound in the region of interest, by using finite number of speakers, is mathematically ill-posed. This problem can be well-posed; in other words, the way to drive the speakers to make desired sound field in the prescribed zone can be directly determined by using energy measure, which is called acoustic brightness, contrast control method. With the method, we can manuver sound ball or balls in space which can generate personal listening zone or virtual sound source. In this paper, a novel way that guides a way to design the speaker array in space to generate sound ball will be introduced. The signals between the zone of interest and the speakers are interpreted as vector spaces. The vector

space interpretation certainly provides insight to design the arrays for acoustic brightness or contrast control. Theoretical formulation in vector space and experimental results of generating the sound ball will be introduced.

11:15

2aSP8. Mode filters and energy conservation. Ilya A. Udovydchenkov (Dept. AOPE, MS#9, Woods Hole Oceanograph. Inst., 266 Woods Hole Rd., Woods Hole, MA 02543), Michael G. Brown (Univ. of Miami, Miami, FL 33149), and Irina I. Rypina (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The discrete form of the mode filtering problem is considered. The relevant equations constitute a linear inverse problem. Solutions to problems of this type, including the mode filtering problem, are subject to a well-known trade-off between resolution and precision. But, unlike the typical linear inverse problem, the correctly formulated mode filtering problem is subject to an energy conservation constraint. This work focuses on the importance of satisfying, approximately at least, the energy conservation constraint when mode filtering is performed. [Work supported by ONR.]

Session 2aUW

Underwater Acoustics, Acoustical Oceanography, and Signal Processing in Acoustics: Acoustic Particle Velocity and Vector Fields: Signal Processing and Communication Applications I

Mohsen Badiey, Cochair
Univ. of Delaware, College of Earth, Ocean, and Environment, Newark, DE 19716-3501

Ali Abdi, Cochair
New Jersey Inst. of Technology, 323 King Blvd., Newark, NJ 07102

Chair's Introduction—8:25

Invited Papers

8:30

2aUW1. Studies of acoustic particle velocity measurements in the ocean. Gerald D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., 291 Rosecrans St., San Diego, CA 92106, gdspain@ucsd.edu)

This presentation reviews previously published results from applying various signal and array processing methods to the analysis of simultaneous measurements of acoustic particle velocity and acoustic pressure in the ocean. Particular focus is placed on interpreting the properties of the elements of the pressure/particle velocity data cross spectral matrix. These properties are used to evaluate measurements of the deep ocean infrasonic noise field during a period of rapidly changing wind speed. Numerical modeling of the evolution in directionality of the active and reactive acoustic intensity vectors during this time period demonstrates the influence of the continental slope on the flow of underwater acoustic energy. [Work supported by the Office of Naval Research.]

8:50

2aUW2. An ultra-low-frequency acoustic vector sensor. Dimitri M. Donskoy (Stevens Inst. of Technol., Hoboken, NJ 07030, ddonskoy@stevens.edu) and Benjamin A. Cray (Naval Undersea Warfare Ctr., Newport, RI 02841)

Passive underwater acoustic surveillance and oceanographic studies, for the most, have not been conducted at ultra-low frequencies (ULFs), from 0.01 Hz to a few hertz. One of the primary reasons for neglecting ULF has been the lack of sensitive, and directional, ULF acoustic sensors. Existing, and many prototyped, vector sensor designs are inertia based and thus not well suited to ULF, typically these sensors that operate at frequencies well above 100 Hz. A new type of acoustic vector sensor, capable of operating within the ULF range and possessing unsurpassed sensitivity (minimal detectable signal levels well below 1 nm/s) was built, tested, and calibrated at the Naval Undersea Warfare Center, Newport, RI. The measurements have validated the proposed detection concept, provided initial performance parameters, and have set forth the foundation for designing the next generation of fieldable prototypes.

9:10

2aUW3. Modeling the acoustic vector field using parabolic equation and normal mode models. Kevin B. Smith (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943)

Numerical algorithms for computing the acoustic vector field, i.e., particle velocity and pressure from an acoustic propagation model, are introduced. Implementation using both a parabolic equation and normal mode approach is considered. The parabolic equation model employed uses a split-step Fourier algorithm, although application of the technique is general to other parabolic equation models. Expressions for the normal mode equations are also presented for both coupled and adiabatic mode models. Results for a Pekeris waveguide are presented for a point source, prompting a brief discussion of multipath influence on the estimation of the direction of energy flow. Approximate analytic solutions are used to validate the general results of both models. Results for the range-dependent benchmark wedge are then presented and show generally good agreement between the two types of models. The results from the two-way, coupled normal mode model provide potential benchmark solutions for the wedge and a means of confirming the accuracy of other models. [Work supported by ONR 321OA.]

9:30

2aUW4. The cooperative array performance experiment: A joint China-US vector field experiment. Daniel Rouseff (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, rouseff@apl.washington.edu), Zhongkang Wang, Shihong Zhou, Hong Meng, and Lisheng Zhou (Hangzhou Appl. Acoust. Res. Inst., Hangzhou City, China)

The cooperative array performance experiment (CAPEx) was performed in Lake Washington near Seattle in September 2009. Acoustic transmissions in the 1.5–4 kHz band were recorded simultaneously on two vertical arrays: one a conventional 32-element pressure-sensor array and the other an 8-element array that measured both pressure and the three orthogonal components of acoustic particle velocity at each element. The present talk is an overview of both the data collected and the hardware used during CAPEx. Data were collected on the stationary arrays for both stationary- and towed-source scenarios in water 60 m deep. The source-receiver range varied between 10 m and 4 km. The data collected at short range demonstrate the relationship between the pressure and particle velocity fields. At more distant ranges, the particle velocity data are used to estimate the bearing to the source. Experimental results are compared to predictions generated using numerical models. [Work supported by ONR.]

9:45

2aUW5. Using the vector sensors as receivers for underwater acoustic communications. T. C. Yang and Alenka Zajić (Naval Res. Lab., Washington, DC 20375, yang@wave.nrl.navy.mil)

A vector sensor package (VSP) generally consists of three orthogonally oriented sensors and an omni directional hydrophone, all packed in a small compact housing. Because of its small size, the VSP might be more useful than a hydrophone array. In contrast to the hydrophone array, which has a non-negligible vertical aperture and is difficult to deploy on an autonomous underwater vehicle (AUV), the VSP can be easily deployed on the AUV and used for underwater acoustic communications (UAC). However, the usefulness of the VSP depends on whether it provides the same spatial diversity as a hydrophone array. Hence, to address this question, this paper investigates the spatial correlation characteristics of the VSP and the bit error rates and/or output SNRs, based on at sea data, between the VSPs and the hydrophone array. The data were collected in May of 2009 on the New Jersey shelf. A hydrophone array and a VSP were deployed close to the ocean bottom and used for UACs. The source was towed at a slow speed, transmitting signals of various modulations at distances 0.5 to a few kilometers from the receivers. Initial results will be presented. [Work supported by the Office of Naval Research.]

10:00—10:30 Break

10:30

2aUW6. Sensitivity analysis of acoustic channel characteristics to sea surface spectral uncertainty. Allan Rosenberg and Qinqing Zhang (11100 Johns Hopkins Rd., Laurel, MD 20723)

The interaction of sound with the sea surface is important for underwater acoustic communications. The high-frequency sound used, 10 kHz and above, is sensitive to surface wave frequencies well above the ~0.5-Hz upper limit routinely measured by wave buoys. Accurately measuring the surface in the short gravity wave regime is difficult and even the general shape of the spectrum is uncertain. The primary measurement challenge is to disentangle the effects of the instrument, including its supporting structure, from what one is attempting to measure. There have been attempts to combine the copious data at low frequencies and the sparse data at higher frequencies to produce model spectra depending on a few parameters that describe the spectrum in the short gravity wave region and above. In this work we study the sensitivity of the acoustic channel characteristics such as the channel impulse response to our uncertain spectral knowledge. We merge low-frequency surface wave spectra measured at NDBC 44014, with various modeled higher-frequency spectra generated from measured environmental parameters to get unified spectra. For each unified spectrum we generate surface realizations, feed them into a rough surface parabolic equation model to compute a channel impulse response, and compare the impulse responses under different surfaces.

10:45

2aUW7. Determination of the location of a sound source in three dimensional based on acoustic vector sensors on the ground. Hans Elias de Bree (Microflown Technologies B.V., P.O. Box 300, 6900 AH Zevenaar, The Netherlands, debree@microflown.com)

An acoustic vector sensor (AVS) consists of three orthogonal particle velocity sensors in combination with a sound pressure microphone. In several publications it has been proven that multiple sources can be located in three dimensions with a single AVS. In this paper it will be shown that it is possible to measure the instantaneous location (this means bearing, elevation, and range) of a single dominant sound source in three dimensional space as well as the angle dependent local ground impedance. Theory as well as results of experiments will be presented.

11:00

2aUW8. A particle velocity gradient beam forming system. Hans-Elias de Bree (Microflown Technologies B.V., P.O. Box 300, 6800AH Zevenaar, The Netherlands, debree@microflown.com)

The topic of this paper is the determination of the acoustic source distribution in the far field with a small, three dimensional (3-D) system consisting closely spaced sound pressure sensitive microphones and particle velocity sensitive Microflown. Sound pressure sensors do have a zero order directionality (that is, no directionality). Particle velocity vector sensors have a first order directionality (this is a cosine shape directionality). With two closely spaced zero order sensors, a first order system can be created. Disadvantages are the low sensitivity for low frequencies and a limited high-frequency response. With two closely spaced first order sensors, a second order system is created. The directionality is a squared cosine shape. With this higher directivity it is possible to create a very small 3-D beam forming system with a reasonable resolution which is the topic of this paper. A second order system can be made with accelerometers or pressure sensors; however, the low-frequency response is very poor. The Microflown has a very high sensitivity at low frequencies so the velocity gradient signal is good. In this paper the velocity gradient method is presented, and a 3-D velocity gradient system is demonstrated.

11:15

2aUW9. An overview of underwater acoustic communication via particle velocity channels: Channel modeling and transceiver design. Ali Abdi (Dept. Elec. & Comput. Eng., New Jersey Inst. of Technol., 323 King Blvd., Newark, NJ, 07102, ali.abdi@njit.edu), Aijun Song, and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

Over the past few decades, the scalar component of the acoustic field, i.e., the pressure channel, has been extensively used for underwater acoustic communication. In recent years, vector components of the acoustic field, such as the three components of acoustic particle velocity, are suggested for underwater communication. Consequently, one can use vector sensors for underwater communication. The small size of vector sensor arrays is an advantage, compared to pressure sensor arrays commonly used in underwater acoustic communication. This is because velocity channels can be measured at a single point in space. So, each vector sensor serves as a multi-channel device. This is particularly useful for compact underwater platforms, such as autonomous underwater vehicles (AUVs). Funded by the National Science Foundation, our research efforts focus on the research problems in two closely related categories: channel modeling and transceiver design. Channel modeling research aims at characterization of those aspects of acoustic particle velocity channels such as delay and Doppler spread, and transmission loss, which determine the communication system performance. Transceiver design addresses optimal use of vector sensors and particle velocity for data modulation and demodulation, equalization, synchronization, and coding. [Work supported by NSF.]

Session 2pAAa

Architectural Acoustics and Speech Communication: Speech Intelligibility and Privacy

Eric L. Reuter, Chair
Reuter Associates, LLC, P.O. Box 4623, Portsmouth, NH 03802-4623

Invited Papers

1:00

2pAAa1. Speech privacy: The new 2010 architectural guidelines. David Sykes (Remington Partners, 23 Buckingham St., Cambridge, MA 02138, dsykes@speechprivacy.org), William Cavanaugh, and Gregory Tocci (Cavanaugh Tocci Assoc., Sudbury, MA)

In January, speech privacy performance criteria were released in the 2010 Guidelines for the Design and Construction of Healthcare Facilities, which is accepted as code by ~42 states and many federal, state, and municipal agencies. This is the first acceptance by regulatory authorities of speech privacy and will impact healthcare facilities across the United States. The criteria have also been adopted as the reference standard for two environmental quality credits in LEED HC and GGHCv2. In addition new 1.5 million dollar penalties took effect on Nov 30, 2009 for violations of the HIPAA Privacy Rule, the first serious penalties since the law was written in 1995. The new speech privacy criteria simplify enforcement by accepting as "equivalent" all four of the established measurement systems (articulation index, privacy index and sound transmission index, and speech intelligibility index), thereby leaving regulators, courts, and professionals with a choice. The developers are working on the next edition (2014) when they hope to introduce the fifth and newest measurement system, S/N(A), which provides specific metrics for both "confidential" and "secure" privacy. Acoustical professionals will need to assist healthcare organizations, regulatory authorities, judges, lawyers, and consumer groups to understand the new criteria.

1:20

2pAAa2. Speech privacy in healthcare: A facility director's and staff. Nikki Rineer (Hope Within Community Health Ctr., 4748 East Harrisburg Pike, Elizabethtown, PA 17022)

Hope Within is a healthcare center that has gone the extra mile to have their architecture spaces evaluated for speech privacy. Once identified they have made accommodations to minimize the risk of overheard conversations and have an ongoing program in place to maintain and improve speech privacy performance. Much has been discussed regarding speech privacy in healthcare, but this paper will discuss the point of view from the director and staff.

1:40

2pAAa3. Measurement of speech privacy of closed rooms using the American Society for Testing and Materials ASTM E2638. Bradford N. Gover and John S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

The degree of speech privacy between various locations in buildings has been of increasing concern, driven in part by legislative requirements and the desire for increased security. Until recently, ASTM E1130 "Standard Test Method for Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index" has been the only standardized measurement method available. This test method is, for a variety of reasons, not suitable for assessing speech privacy of closed rooms or conditions of very high-speech privacy. These issues are, however, addressed by the new ASTM E2638 "Standard Test Method for Objective Measurement of the Speech Privacy Provided by a Closed Room." ASTM E2638 describes a test method suitable for enclosed rooms of nearly all sizes and defines a measure called speech privacy class that can be used to accurately rate speech privacy over a very broad range, from no privacy to very high secrecy. The new E2638 measurement method will be described, and measurements in several real rooms will be presented.

2:00

2pAAa4. Setting speech privacy criteria for closed rooms in terms of speech privacy class values. John S. Bradley and Bradford N. Gover (Natl. Res. Council, 1200 Montreal Rd., Ottawa, Canada)

The new ASTM E2638 standard defines speech privacy class (SPC) as the sum of the measured average noise level at the position of a potential eavesdropper outside the room and the measured level difference between a source room average and the transmitted levels at the same potential eavesdropper location. For a given situation, the likelihood of transmitted speech being audible or intelligible can be related to the probability of higher-speech levels occurring in the meeting room. Increasing speech privacy criteria can be defined in terms of increasing SPC values. For a particular meeting room speech level, there is an SPC value for which transmitted speech would be below the threshold of intelligibility or even below the threshold of audibility. One can therefore create a set of increasing SPC values corresponding to increasing speech privacy and for each SPC value, one can give the probability of transmitted speech being either audible or intelligible. This paper describes a new procedure that makes it possible to specify degrees of speech privacy for meeting rooms and offices in terms of the expected risk of a privacy lapse and varying from quite minimal to extremely high-speech privacy.

2:20

2pAAa5. Low background noise is needed for good student-to-student classroom communication. David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683, dlubman@dlacoustics.com)

ANSI standard S12.60 requires that background noise not exceed 35 dBA in unoccupied lecture classrooms. Despite the strong and broad research basis for ANSI's 35-dBA requirement, some classroom noise guidelines permit 45 dBA (notably California Collaborative for High Performance Schools and Leadership in Energy and Environmental Design). How did this come about? The 45-dBA limit was a compromise between the 50-dBA guidelines then prevailing and ANSI's 35-dBA limits. There is no audiological basis for a 45-dBA BNL limit. With 45-dBA BNLs, the speech to noise ratio (SNR) in the rear of small classrooms is well below the 15-dB minimum required for sustained learning. With 35-dBA BNLs, the SNR approaches ANSI's 15-dB minimum. It is underappreciated that BNLs of 35 dBA or less also promote better student-to-student (STS) communication. For every 10 dB of background noise reduction, the physical range of STS communication increases by a factor of about 3.16, and the floor area encompassed increases by a factor of 10. As BNL is reduced, eventually all classroom occupants are within communication range.

2:40

2pAAa6. A primer on classroom acoustics. Gary W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611), Reece Skelton, Keely M. Siebein, and Chris P. Jones (Siebein Assoc., Inc., Gainesville, FL 32607)

Research involving quantitative and qualitative evaluation of classroom activities has shown that speech perception and intelligibility in elementary, middle, and high schools can be measured, modeled, and simulated based on an impulse response-based theory. The theory combines the vocal production of the speaker, the hearing abilities of the listeners, the strength of the direct sound and early sound reflections from room surfaces, duration of longer term reverberation, and the level, type, information content, and frequency content of intruding sounds. Regression models relating measured intelligibility in classrooms to measured, modeled, and simulated sounds show model $r2 > 0.90$. A model relating speech perception score to the level, type, information content, and frequency content of typical classroom noise reaches $r2 > 0.90$ as well. The impulse response based measures are reduced to a series of architectural systems that can be implemented using alternate construction systems to meet budget requirements.

3:00—3:15 Break

3:15

2pAAa7. Assessing the potential intelligibility of assistive listening systems. Peter Mapp (Peter Mapp Assoc., Colchester, United Kingdom, peter@petermapp.com)

Approximately 14% of the general population suffer from a noticeable degree of hearing loss and would benefit from some form of hearing assistance or deaf aid. Recent DDA legislation and requirements mean that many more assistive listening systems are being installed, yet there is evidence to suggest that many of these systems fail to perform adequately and provide the benefit expected. There has also been a proliferation of classroom and lecture room "soundfield" systems, with much conflicting evidence as to their apparent effectiveness. This paper reports on the results of some trial acoustic performance testing of such systems. In particular, the effects of system microphone type, distance, and location are shown to have a significant effect on the resultant performance. The potential of using the sound transmission index and, in particular, STIPa for carrying out installation surveys has been investigated, and a number of practical problems are highlighted. The requirements for a suitable acoustic test source to mimic a human talker are discussed. The findings discussed in the paper are also relevant to the installation and testing of "soundfield systems."

3:35

2pAAa8. Effects of noise type and signal-to-noise ratio on speech intelligibility and the speech transmission index. Stephen D. Secules (Arup Acoust., 13 Fitzroy St., London W1T 4BQ, United Kingdom)

The STI metric is the most reliable measure we have for assessing the intelligibility of speech in public address/voice alarm or speech reinforcement systems. A good design for speech is based on an understanding of the room acoustics and electroacoustics of the system, with a consideration for the expected signal to noise ratios in the room. In modeling, most noise signals are assumed to be broadband, whereas in reality the noise signals can be significantly variable in the frequency and time domain, which may dramatically increase or reduce the intelligibility. This study compares the effects of type of noise on intelligibility as measured through both word scores and STI. It explores masking from both theoretical and real world signals. The determined relationships will help inform the design process by clarifying the effects of non-broadband and variable noise signals on intelligibility.

Contributed Papers

3:55

2pAAa9. Comparing head rotation angle, visual localization, and recall-proficiency of school-aged children 8–12 while listening to a story by multiple discreet talkers in a virtual classroom. Daniel L. Valente, Dawna E. Lewis, John M. Franco, and Elizabeth C. Heinrichs (Dept. of Res., Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

In general, children perform more poorly in speech intelligibility tasks than adults while in a noisy environment such as a classroom. This is especially true in discussion situations where the active talkers often are located to the side or behind the student. The purpose of this study was to obtain baseline data for normal hearing (NH) students who experienced multiple-talker stimuli in a controlled virtual classroom environment. Data were gathered using a gyroscopic head-tracker and a post-story comprehension task. Twenty elementary-aged students (ages 8–12, four students per year) were positioned in the center of the classroom and presented a story read by five talkers positioned around the student (reproduced by means of loudspeakers and LCD monitors). Students used different strategies in terms of head rotation over time while performing the experiment and individual differences were seen both across subjects within an age group and between age groups. The post-test comprehension task was used as a metric of performance in the environment. Results from the NH students are compared to a control group of adult NH listeners using head rotation angle over time, comprehension scores, and localization accuracy as salient metrics.

4:10

2pAAa10. Verifying speech intelligibility in classrooms. Roman Wowk (Papadimos Group, 818 Fifth St., Ste. 207, San Rafael, CA 94901)

Classroom speech intelligibility has undoubtedly become a hot topic and is capturing attention on all fronts. While there has been widespread adoption of many new guidelines and standards, even the most precise initiatives will not reach their full potential if speech intelligibility is not directly verified in classrooms after they are built. Performance verification ultimately encourages a better final product and provides crucial measured data for validating design methods. To ensure measurable performance, criteria would have to be established at the outset of the design using a metric directly related to speech intelligibility such as speech transmission index. Current guidelines generally focus on reverberation time (RT) and background noise for improving speech communication in classrooms. While low background noise is a necessary component for speech intelligibility, there is not always a direct correlation between RT and speech intelligibility. For instance, improper placement of absorptive treatments for reverberation control could eliminate early supporting reflections which are beneficial for speech intelligibility. With computer modeling becoming widespread and relatively inexpensive, designing directly for speech intelligibility is becoming more practical. This paper explores case studies and examines the overall process for optimizing speech communication in classrooms with an emphasis on performance verification.

4:25

2pAAa11. Ambient noise levels and reverberation times in Mississippi school rooms. Edward L. Goshorn, Megan N. Lucus, and Brett E. Kemker (Dept. of Speech and Hearing Sci., Psychoacoustics Lab., Univ. of Southern Mississippi, 118, College Dr. #5092, Hattiesburg, MS 39406)

Nine elementary school classrooms at three Mississippi public schools were selected at random for noise and reverberation time measures to monitor voluntary compliance with ANSI standard S12.60-2002 (Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools). 1-h equivalent dBA noise levels and reverberation times (T60) were measured with SPECTRAPLUS Version 5.0 software. Measures were taken on separate days at each school in unoccupied classrooms with the HVAC system operating. For ambient noise measures, an Audix TR40 omnidirectional microphone was placed in a 12 in. stand on a student's desk near the middle of the room and oriented toward the teacher's desk at a 45 deg angle. For the reverberation time measure a 400 W loudspeaker (Yamaha BR15M) was placed on the teacher's desk about 3 m from the microphone. Results showed that noise levels varied from 28 to 51 dBA with broad spectra containing peaks at 120 Hz and harmonics. Five of nine classrooms had noise levels that exceeded the ANSI recommended level of 35 dBA. Reverberation times varied from 0.21 to 0.62 s; only one classroom exceeded the recommended minimal reverberation time of 0.6 s. Acoustical modifications to reduce noise levels and reverberation times will be addressed.

4:40

2pAAa12. Voice privacy associated with raised floor law office: A case study. Christopher J. Pollock and Geoffrey Sparks (Shen Milsom & Wilke, LLC, 3300 N Fairfax Dr., Ste. 302, Arlington VA 22201, cpollock@smwllc.com)

Presentation and discussion of the acoustical transmission and voice privacy issues associated with raised floor office fit out construction. Many new LEED office fit out projects are using raised floor constructions for future flexibilty and use of underfloor plenum for HVAC systems and/or technology cable plant. Issues of user expectations of privacy versus capabilities of the system are becoming more common. Case study office fit out was completed with raised floor and the extent of voice transmission due to various construction conditions limited voice privacy and created noise disruption from offices, workrooms, and conference rooms. Case study involves a law office where privacy concerns are heightened due to sensitive discussions. Measurements of noise isolation and privacy were made both before and after modifications, and results will be presented and discussed and conclusions for practitioners will be discussed.

TUESDAY AFTERNOON, 20 APRIL 2010 1:00 TO 3:00 P.M.
GRAND BALLROOM FOYER WEST/NORTH

Session 2pAAb

Architectural Acoustics and National Council of Acoustical Consulants: Student Design Competition

Andrew Miller, Cochair
BAi, LLC, 4006 Speedway, Austin, TX 78751

Byron Harrison, Cochair
The Talaske Group Inc., 1033 South Blvd., Ste. 101, Oak Park, IL 60302

Robert C. Coffeen, Cochair
University of Kansas, School of Architecture and Urban Design, Marvin Hall, Lawrence, KS 66045

.

The Technical Committee on Architectural Acoustics of the Acoustical Society of America, the Robert Bradford Newman Student Award Fund, and the National Council of Acoustical Consultants are sponsoring the 2010 Student Design Competition that will be professionally judged at this meeting. The purpose of this design competition is to encourage students enrolled in architecture, engineering, physics, and other university curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics in the design of performance hall primarily for opera on a college campus of moderate size.

This competition is open to undergraduate and graduate students from all nations. Submissions will be poster presentations that demonstrate room acoustics, noise control, and acoustic isolation techniques in building planning and room design. The submitted designs will be displayed in this session and they will be judged by a panel of professional architects and acoustical consultants. An award of $1250.00 US will be made to the entry judged "First Honors." Four awards of $700.00 US will be made to each of four entries judged "Commendation."

Session 2pAB

Animal Bioacoustics: Estimating Spatial Density of Animal Populations with Passive Acoustics

David K. Mellinger, Chair
Oregon State Univ., 2030 S.E. Marine Science Dr., Newport, OR 97365

Chair's Introduction—1:00

Invited Papers

1:05

2pAB1. Minke whale boing vocalization density estimation at the Pacific Missile Range Facility, Hawaii. Stephen W. Martin (Space and Naval Warfare Systems Ctr. Pacific, 53366 Front St., San Diego, CA 92152), Len Thomas, Tiago A. Marques (Univ. of St. Andrews, St. Andrews KY16 9LZ, Scotland), Ronald P. Morrissey, Susan Jarvis, Nancy DiMarzio (Naval Undersea Warfare Ctr., Newport, RI 02841), and David K. Mellinger (Oregon St. Univ. and NOAA Pacific Marine Environ. Lab., Newport, OR 97365)

Minke whales (*Balaenoptera acutorostrata*) seasonally visit the Hawaiian Islands with an apparent peak in activity in the February /March timeframe. Their presence is readily detected acoustically from their boing vocalizations, which are detectable on the Pacific Missile Range Facilities (PMRF) deep-water (3500 m to over 4500-m depth) bottom mounted hydrophones northwest of Kauai, HI. Automated techniques were developed to detect, classify, and associate the boings across 16 separate hydrophones. Over 13 h of data for each of 12 separate days from the 2006 whale season (February/March/April) were processed for detection and classification (D/C) of boing vocalizations. A sample set of the data (10 min from each day) was analyzed to characterize the detection function (probability of detection as a function of horizontal range) and the false positive rate of the boing D/C. Boings were associated across hydrophones, defining a capture history for each boing vocalization. Spatially explicit capture recapture methods were applied to the sample data associations to derive the detection function. Applying the parameters estimated on the sample data set toward the full data ser allows an estimation of the minke whale boing vocalization density in this area over these time periods in 2006.

1:25

2pAB2. Estimating density from single hydrophones by means of propagation modeling. Elizabeth T. Küsel, David K. Mellinger (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 S. Marine Sci. Dr., Newport, OR 97365, elizabeth.kusel@oregonstate.edu), Len Thomas, Tiago A. Marques (Univ. of St. Andrews, Fife KY16 9AJ, Scotland), David J. Moretti, and Jessica Ward (Naval Undersea Warfare Ctr., Newport, RI 02841)

A density estimation method developed under the project Density Estimation for Cetaceans from passive Acoustic Fixed sensors is presented. It uses sound propagation modeling to estimate the probability of detecting an animal as a function of its distance from the receiving sensor. Two case studies involving Blainville's beaked whales (*Mesoplodon densirostris*) and sperm whales (*Physeter macrocephalus*) are analyzed. The study area is the Atlantic Undersea Test and Evaluation Center in the Tongue of the Ocean, Bahamas. A ray-tracing acoustic propagation model is used to estimate the environmental transmission loss as a function of depth and range in several directions away from a single hydrophone. The computed transmission loss is compared to ambient noise levels, source level, and beam pattern distributions available in the literature to estimate detection probability as a function of range. Detection threshold is characterized from the signal-to-noise ratio of detected clicks. Information on click production rate is also taken into account by the density estimation model. Beaked whale detection probability function provides a relevant comparison to both the detection function and the spatial density of whales derived empirically from the DTag data by Marques *et al.* [(2009)].

1:45

2pAB3. Population density of sperm whales in the Bahamas estimated using propagation modeling. David K. Mellinger, Elizabeth T. Küsel (Coop. Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, david.mellinger@oregonstate.edu), Len Thomas, Tiago Marques (Univ. of St Andrews, Scotland), David Moretti, Paul Baggenstoss, Jessica Ward, Nancy DiMarzio, and Ron Morrissey (Naval Undersea Warfare Ctr., Newport, RI)

The population density of sperm whales is estimated using an acoustic model to calculate the probability of receiving their clicks. The model uses estimates of (1) the source level of clicks, (2) the beampattern of the whales' emitted clicks, (3) the distribution of whale orientations, (4) the loss between source and receiver (derived from acoustic propagation modeling), (5) noise levels at the receiver, (6) the detector's rates of missed calls and false detections, and (7) sperm whales' average click rate. These data are combined in a model that propagates simulated clicks from whales at various simulated positions to the receiving hydrophone to estimate the detection function—the probability of receiving a click as a function of distance. This function is then combined with information on whale click rates (performed at a chosen time of day when sperm whales appear to click at a measurably predictable rate) to estimate the population density corresponding to the number of received clicks in a given period of time. Data from the U.S. Navy's Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas are used to estimate sperm whale population density there. [Thanks to ONR, NOAA, and the Navy's Environmental Readiness Division for funding.]

2:05

2pAB4. Density estimation of leopard seals using a single stationary passive acoustic sensor. Holger Klinck (Hatfield Marine Sci. Ctr., Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, holger.klinck@oregonstate.edu), Nadine Constantinou (Univ. of New South Wales, Sydney, New South Wales 2052, Australia), David K. Mellinger (Oregon State Univ., Newport, OR 97365), and Tracey Rogers (Univ. of New South Wales, Sydney, New South Wales 2052, Australia)

The objective of this study is to estimate the spatial density of leopard seals using data recorded with a single stationary passive acoustic recording system in the Bransfield Strait, Antarctica, between 2005 and 2007. The most prominent vocalization of the leopard seal—the low double trill (LDT)—is used as a proxy for the presence of the species in the vicinity of the recording system. Because of the stereotypic nature and high frequency of occurrence of the LDT, a long-term spectrogram approach can be applied to the data sets to reliably detect the presence of the target species. Energy levels in the target frequency band (200–400 Hz) as derived by the long-term spectrogram analysis are related to number of manually counted calls extracted for selected periods. A linear regression analysis showed that energy levels are highly correlated with the number of manually counted calls. The number of recorded calls per unit time is converted into number of vocalizing animals per unit time by applying published calling rates for this species. In a last step of the analysis, the detection area is defined and leopard seal densities estimated. Opportunities and challenges of the method will be discussed.

2:25

2pAB5. Determining the spatial distribution of an Antarctic top predator using passive acoustics. Nadine E. Constantinou, Tracey L. Rogers (Evolution and Ecology Res. Ctr. (EERC), School of Biological, Earth and Environ. Sci. (BEES), Univ. of New South Wales (UNSW), Sydney, New South Wales 2052, Australia, n.constantinou@student.unsw.edu.au), Shawn W. Laffan, and David I. Warton (UNSW, New South Wales 2052, Australia)

The leopard seal (*Hydrurga leptonyx*) is one of four species of ice seals in Antarctica with each species occupying a distinct position in the Antarctic sea-ice ecosystem. Ice seals offer a potential source of information about ecosystem interactions and environmental variability integrated over a variety of spatial and temporal scales. During the austral spring and summer, leopard seals move within the pack ice to breed. Acoustic surveying is necessary to assess their distributions as male leopard seals vocalize underwater as part of their breeding display. During the 1999/2000 austral summer, the relative abundance of adult male leopard seals was determined using underwater passive acoustic point-transect surveys. The abundance data were combined with environmental data in a geographical information system, and a model was developed to determine what factors of the environment are correlated with their abundance and distribution. The model with the best predictive power showed a trend of increased abundance toward the pack ice edge. These regions are associated with areas of increased foraging potential suggesting that the distribution of leopard seals off Eastern Antarctica is influenced, in part, by increased availability of prey.

2:45—3:05 Break

Contributed Papers

3:05

2pAB6. Combining visual and fixed passive acoustic methods to measure annual variability of cetacean occurrence at the NE-coast of Iceland. Edda E. Magnusdottir, Marianne H. Rasmussen (Husavik Res. Ctr., Univ. of Iceland, Hafnarstett 3, 640 Husavik, eem@hi.is), and Marc Lammers (Hawaii Inst. of Marine Biology, Kailua, HI 96834)

In order to measure cetacean occurrence at the NE-coast of Iceland, two seabed-mounted Ecological Acoustic Recorders (EARs) were deployed in Skjálfandi Bay in September 2008, deployments will continue throughout September 2010. To obtain species confirmation, visual observations were conducted from a lighthouse in October 2008 to September 2009 with total observation effort of 202 h. With visual observations, the aim was to acquire background information on the cetacean sighted in the area to improve the estimates of occurrence in continued long-term passive acoustic recordings in Icelandic and adjacent waters. Movements and locations of the animals were recorded using a theodolite and a real-time mapping program (CYCLOPS TRACKER). Visual observations included sightings of blue whales (*Balenoptera musculus*), Humpback whales (*Megaptera novaeangliae*), Minke whales (*Balaenoptera acutorostrata*), white-beaked dolphins (*Lagernorhynchus albirostris*) and killer whales (*Orcinus orca*). A custom MATLAB program is used to automatically detect and analyze target biological sounds produced by cetaceans or by pelagic fish recorded on the EARs. The visual data are compared to the occurrence of identified sounds in order to better understand the vocal activity and average silence periods of the cetaceans. Also, detected vocalizations are compared to the number, location, and behavior of simultaneously sighted animals.

3:20

2pAB7. Seasonal and diurnal presence of finless porpoises at the corridor to the ocean from a semi-closed bay habitat. Tomonari Akamatsu (NRIFE, FRA, Hasaki, Kamisu, Ibaraki 314-0408, Japan), Kiyomi Nakamura (Kujyu Kushima Aquarium, Sasebo, Nagasaki 858-0922, Japan), Ryo Kawabe, Seishiro Furukawa, Hiromi Murata (Nagasaki Univ., Nagasaki 851-2213, Japan), Akihiro Kawakubo, and Masayuki Komaba (Kujyu Kushima Aquarium, Sasebo, Nagasaki 858-0922, Japan)

Finless porpoises in Omura Bay is the smallest among five populations in Japanese waters. This is a newly established population after the global warming of 9000 years ago, consisting of approximately 300 individuals and believed to be confined in Omura Bay. At the major corridor to the ocean from this bay, presence of finless porpoises was monitored acoustically from November 2007 to May 2009. A stereo acoustic event recorder (A-tag) was deployed at this strait. A-tag stored the intensity and the sound source direction of biosonar signals that provides independent traces of sound source corresponding to each animal. During 1.5 year observation, 226 individuals of porpoises were detected. Of them, 76% were presented at night and 73% observed during March to April. In the same season, anchovy landings in the Omura bay was recorded. 70% percent of the porpoises were observed during the tidal current that went out of Omura Bay that suggests the porpoises swam back and forth around the monitoring station. Although finless porpoises have been known to be confined in a local area, occasional migration of this species induced by prey availability might extend their habitat at different waters in Asian coasts.

3:35

2pAB8. Are passive acoustics an appropriate method to accurately represent a soniferous Atlantic croaker (*Micropogonias undulatus*) population? Cecilia S. Krahforst, Joseph J. Luczkovich, and Mark W. Sprague (Dept. of Biology, East Carolina Univ., Greenville, NC 27858, csk1021@ecu.edu)

It was observed that the fundamental frequency of Atlantic croaker sounds is inversely correlated with the length of the fish. We used fundamental frequency as a method to estimate the average length of the croaker population at any given time. Croaker were collected using an otter trawl and simultaneously counted and measured acoustically with a 200-kHz split-beam echosounder at two sites within the Pamlico Sound estuary, NC from June to November 2008. Passive acoustic recorders [long-term acoustic recording system (LARS) recordings 10-s wave files $<$ 10 kHz at 15-min intervals] were deployed near each trawling site to obtain *in-situ* recordings of croaker over the same period. Based on captive fish recordings, a linear regression analysis related total length to fundamental frequency, where TL = 305.323 mm$-$(0.270 Hz^{-1})f_0(r^2 = 0.84). This equation was then used to estimate croaker lengths from LARS recordings. Lengths of fish collected in the trawls, compared with the estimate from the passive LARS recordings and the echosounder surveys, were significantly smaller ($P > 0.0001$). Thus suggests that size-selectivity of the trawl underestimates the average size of fish and that acoustic methods (active and passive) provide a more accurate size estimate for croaker populations.

TUESDAY AFTERNOON, 20 APRIL 2010 KENT A/B, 1:00 TO 5:00 P.M.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Numerical Modeling for Medical Ultrasound I

Robert J. McGough, Cochair
Michigan State Univ., Dept. of Electrical and Computer Engineering, 2120 Engineering Bldg., East Lansing, MI 48824

Vera A. Khokhlova, Cochair
Moscow State Univ., Acoustics Dept., Leninskie Gory, Moscow, 119992, Russia

Chair's Introduction—1:00

Invited Papers

1:05

2pBB1. A parallel algorithm for high-intensity focused ultrasound simulation. Joshua E. Soneson (Div. of Solid and Fluid Mech., U.S. Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993, joshua.soneson@fda.hhs.gov)

The Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation serves as today's workhorse for the modeling and simulation of high-intensity focused ultrasound. This approximation of the full nonlinear wave equation allows serial algorithms to compute the pressure distribution produced by circular transducers usually in several hours on a modern workstation. However, more complicated transducer geometries preclude axisymmetric modeling, and extending into the third spatial dimension requires significantly more computational power, making serial algorithms intractable. This presentation details a parallel implementation of a combination time and frequency domain algorithm for rapid solution of the KZK equation. A split step method is used in which the linear terms in KZK are solved in the frequency domain and the nonlinear term is solved in the time domain. In the frequency domain, each harmonic may be computed independently, so this task is distributed over the available processors. In the time domain, the transverse spatial domain may be partitioned and distributed for independent solution of the nonlinear term. Near-linear scaling (acceleration vs number of processors) using this algorithm is demonstrated.

1:25

2pBB2. Analytic and numerical modeling of ultrasonic B-scan and echo decorrelation imaging. T. Douglas Mast and Swetha Subramanian (Dept. of Biomedical Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586, doug.mast@uc.edu)

An numerical model is presented for *B*-scan images of weakly scattering, lossy media, based on ultrasound array beam patterns calculated analytically under the Fresnel approximation. Given these beam patterns and a three-dimensional analytic or numerical tissue model, this method yields beamformed *A*-line signals from which *B*-scan images are constructed. This approach is further employed to model echo decorrelation imaging, a method for quantitatively mapping transient heat-induced changes in pulse-echo ultrasound images. In echo decorrelation imaging, a normalized decorrelation parameter is computed between *A*-line signals separated by milliseconds. Maps of this parameter comprise echo decorrelation images, which are potentially useful for monitoring of local tissue coagulation during thermal ablation treatments for cancer therapy. Following previous studies in which scattering cross section has been related to spatial-frequency spectra of tissue sound speed, density, and impedance variations, echo decorrelation is related quantitatively to the local decoherence of these spatial-frequency spectra. Decoherence estimates are validated by simulations employing analytic array beam patterns and random-media models for ablated tissue and are further applied to quantify tissue structure changes caused by thermal coagulation during radiofrequency ablation of *ex-vivo* bovine liver tissue.

1:45

2pBB3. A modeling tool for therapeutic and imaging ultrasound applications. Francesco Curra (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The fields of therapeutic and imaging ultrasound have broad medical applications. However, the inherent complexity of biological media and the nonlinear nature of ultrasound propagation at therapeutic regimes make optimization and control of the therapy still a challenging task. An accurate modeling tool for solving multi-dimensional ultrasound-based problems in complex geometries that can greatly assist in the optimization and control of the treatment is presented. The model consists of 2-D, 2.5-D (cylindrical symmetry), and 3-D coupled solutions for acoustic and elastic wave propagation in heterogeneous, lossy media complemented by the bioheat transfer equation for temperature estimation. It includes linear and nonlinear wave propagations, arbitrary frequency power law for attenuation, and can account for multiple reflections and backscattered fields. Work is underway for the inclusion of cavitation effects and the extension of the model to adaptive grid refinement and unstructured, deformable grids. Sample results for therapeutic and imaging applications will be presented. [Work supported by US Army MRMC and NIH NIDDK.]

2:05

2pBB4. Modeling of high-frequency ultrasound contrast agents. John Allen, Fanny Cugnet (Dept. of Mech. Eng., Univ. of Hawaii-Manoa, 2540 Dole St., Honolulu, HI 96822, alleniii@hawaii.edu), Jonathan Mamou, Paul Lee, Parag Chitnis, and Jeff Ketterling (Riverside Res. Inst., New York, NY 10038)

High-frequency applications of ultrasound contrast agents continue to develop with their increasing use for imaging intravascular locations as well as dermal and ocular tissue. However, conventional ultrasound contrast agents were originally designed for traditionally lower diagnostic frequencies. The optimal contrast agent design and related acoustic forcing methods for high-frequency applications are important topics of on-going research. In this respect,ultrasound contrast agent models are reviewed and, in particular, the role of the shell is highlighted for novel high-frequency applications. Shell material and related stability issues are discussed in terms of the constitutive equation assumptions. The concept of auto-resonance, which may be achieved by a slowly varying forcing frequency, is introduced as a method for enhancing the subharmonic response of free or bound agents. The phase locking characteristics of auto-resonance are analyzed using an empirical mode decomposition of the scattered signal. This provides the intrinsic mode functions from which the associated instantaneous frequencies and phases can be determined. The Shannon entropy and the variance of the phase are used to quantify the behavior. Comparisons of the numerical results are made with current experimental data. [*Support from* NIH-EB006372 and NIH-2G12-RR003016121.]

2:25

2pBB5. Modeling of nonlinear shock wave propagation and thermal effects in high-intensity focused ultrasound fields. Vera A. Khokhlova (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105 and Faculty of Phys., Moscow State Univ., Moscow 119991, Russia, vera@acs366.phys.msu.ru), Olga V. Bessonova, Mikhail V. Averiyanov (Moscow State Univ., Moscow 119991, Russia), Joshua E. Soneson (U.S. Food and Drug Administration, Silver Spring, MD 20993), and Robin O. Cleveland (Boston Univ., Boston, MA 02215)

Numerical simulations based on the Khokhlov–Zabolotskaya-type equation are currently used to characterize therapeutic high-intensity focused ultrasound fields in water and to predict bioeffects in tissue. Here results from three different algorithms that differ in calculating the nonlinear term in the equation are presented. Shock capturing schemes of Godunov type, exact implicit solution with further extrapolation of the waveform over a uniform temporal grid, and direct modeling in the frequency domain are tested. In the case of weak nonlinearity, all schemes give essentially the same solution. However, at high peak pressures around 50 MPa and strong shocks developed in the focal region, the predictions of acoustic variables and heat deposition become sensitive to the algorithm employed. The parameters of the schemes, such as number of harmonics or temporal samples and the inclusion of artificial absorption that provides consistent results, are discussed. It is shown that the spectral and Godunov-type approaches require about 6 points and implicit time domain approach needs more than 50 points in the shock to be accurate. In all schemes artificial absorption should be employed to obtain acceptable accuracy with fewer points per cycle. [Work supported by the NIH EB007643, ISTC 3691, RFBR 09-02-01530, and NSF 0835795 grants.]

2:45

2pBB6. Generalization of the iterative nonlinear contrast source method to realistic, nonlinear biomedical tissue. Martin D. Verweij (Lab. of Electromagnetic Res., Fac. of Elec. Eng., Mathematics and Comput. Sci., Delft Univ. of Technol., Mekelweg 4, 2628 CD Delft, The Netherlands)

Originally, the iterative nonlinear contrast source (INCS) method has been developed to compute the nonlinear, wide-angle, pulsed ultrasound field in a homogeneous and lossless medium. The method considers the nonlinear term of the lossless Westervelt equation as a distributed contrast source in a linear background medium. The full nonlinear wave field follows from the Neumann iterative solution of the resulting integral equation. Each iteration step involves the spatiotemporal convolution of the background Green's function with an estimate of the contrast source. Appropriate filtering provides accurate field predictions for a discretization approaching two points per smallest wavelength or period. The current paper first discusses the theoretical generalization of the original contrast source approach, e.g., to include tissue attenuation or inhomogeneity. Next, computational results are presented that show the directional independence (even for the nonlinear distortion) and the wide-angle performance of the original INCS method. Finally, results are presented for extensions that can deal with power law losses or inhomogeneity in the wave speed. The wide-angle capability and the versatility of the contrast source approach demonstrate that the INCS method is well-suited for modeling the various aspects of realistic biomedical tissue behavior. [Work supported by STW and NCF.]

3:05—3:35 Break

3:35

2pBB7. Fractional wave equation for compressional and shear waves. S. Holm (Dept. Informatics, Univ. of Oslo, P.O. Box 1080, NO-0316 Oslo, Norway) and R. Sinkus (Inst. Langevin, ESPCI ParisTech, F-75231 Paris Cedex 05, France)

This study has been motivated by the observed difference in the range of the power law attenuation exponent for compressional and shear waves. Usually compressional attenuation increases with frequency to a power between 1 and 2, while shear wave attenuation often is described with powers less than 1. Another motivation is the apparent lack of partial differential equations with desirable properties such as causality that describe such wave propagation. Starting with a constitutive equation which is generalized Hooke's law with a loss term containing a fractional derivative, one can derive a causal fractional wave equation previously given by Caputo (1967) and Wismer (2006). In the low $\omega\tau$ (low-frequency) case, this equation has an attenuation with a power law in the range from 1 to 2. This is consistent with, e.g., attenuation in tissue. In the often neglected high $\omega\tau$ (high-frequency) case, it describes attenuation with a power law between 0 and 1, consistent with what is observed in, e.g., dynamic elastography. Thus a unifying wave equation derived properly from constitutive equations can describe both cases. Recent results on fractional random media will also be discussed as they give similar characteristics from a microperspective.

3:55

2pBB8. Transducer models in the ultrasound simulation program FIELD II and their accuracy. J. A. Jensen and D. Bk (Dept. of Biomedical Eng., Ctr. for Fast Ultrasound Imaging, Tech. Univ. of Denmark, Build. 349, DK-2800 Lyngby, Denmark, jaj@elektro.dtu.dk)

The FIELD II simulation program can be used for simulating any kind of linear ultrasound fields. The program is capable of describing multi-element transducers used with any kind of excitation, apodization, and focusing. The program has been widely used in both academia and by commercial ultrasound companies for investigation novel transducer geometries and advanced linear imaging schemes. The program models transducer geometries using a division of the transducer elements into either rectangles, triangles, or bounding lines. The precision of the simulation and the simulation time is intimately linked through the choice of the fundamental elements. The rectangular elements use a far-field approximation, whereas the two other methods use the full analytic solution, leading to a higher precision at the price of a slower simulation time. The talk will describe the different compromises and solutions to obtain a fast simulation and still attain a high precision including a newly developed semi-analytic solution for a convex surface elements.

Contributed Papers

4:15

2pBB9. Beamforming for vibro-acoustography using a 1.75D array transducer. Matthew W. Urban, Mostafa Fatemi, and Azra Alizad (Dept. of Physio. and Biomedical Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, urban.matthew@mayo.edu)

Vibro-acoustography is an ultrasound-based imaging modality that maps the acoustic response of an object induced by acoustic radiation force. The method employs two ultrasound beams with frequencies separated by a small difference, typically in the kilohertz range. Using a 1.75D array transducer, which has multiple rows of elements, offers an advantage over linear array transducers because it allows for focusing in the elevation direction. We have performed simulations of different arrangements for the apertures assigned to the two ultrasound frequencies and evaluated resolution and sidelobe levels for different arrangements. However, the large size of the elements produces grating lobes in the transmitted fields. We will describe metrics for evaluating grating lobe levels compared to the main lobe of the transmitted field as a quantitative approach to optimize aperture design. We present experimental measurements of the transmitted field and compare with simulation results. Finally, we present imaging results from phantom experiments. This beamforming study is directed toward improving vibro-acoustography image formation using a General Electric Vivid 7 scanner for breast and thyroid imaging. [Mayo Clinic and two of the authors have potential financial interest related to devices or technology referenced above. Work supported by NIH Grant No. R21CA121579 and AIUM Award No. AIUM-EER#1.]

4:30

2pBB10. Calculations of intensities for radiation force modeling with the commercial software package FOCUS. Robert J. McGough, Donald J. Vanderlaan, Alexander Dutch (Dept. of Elec. and Comput. Eng., Michigan State Univ., East Lansing, MI), and Matthew W. Urban (Mayo Clinic College of Medicine, Rochester, MN)

Rapid calculations of the intensity are required for models of the radiation force generated by linear ultrasound phased arrays. To determine the computation times and numerical errors produced by two medical ultrasound simulation packages, namely, FIELD II and FOCUS, comparisons of the calculated pressures and intensities are performed using a simulated Vermon L5 phased array probe with 128 active elements. In these simulations, the array generates a fixed focus 25 mm from the array for a 5-MHz excitation, and the pressure field is computed on a $133 \times 133 \times 261$ rectilinear grid that is centered at the focus. Half wavelength sampling is employed in all three directions. In FIELD II, this calculation takes approximately 4 h on a 64-bit laptop running Linux, and in FOCUS, the same calculation takes less than 4 min on the same computer. The numerical errors are also much smaller with FOCUS. Details of the simulation approaches employed for this transducer model will be presented, including selection of FIELD II and FOCUS simulation parameters and calculations of the reference pressure field for error evaluations. The impact of numerical errors on shear wave calculations will also be discussed.

4:45

2pBB11. Strategies to improve conformal tumor heating with waveform diversity beamforming. Robert J. McGough and Matthew R. Jennings (Dept. of Elec. and Comput. Eng., Michigan State Univ., East Lansing, MI)

Waveform diversity is a phased array beamforming approach that, when applied to tumor heating with ultrasound, simultaneously maximizes the power at control points in the tumor while minimizing the power deposited in specified sensitive tissues. When waveform diversity calculations are combined with a rank deficient beamforming method, the optimal number of focal patterns is determined, and the optimal array excitations are also obtained. Previous results have shown that waveform diversity combined with mode scanning reduces problems with intervening tissue heating that otherwise occur when standard spot scanning approaches are applied. In an effort to extend these results and achieve a more conformal tumor heat deposition, variable weights will be applied to each of the focal points as a function of distance from the center of the spherical tumor volume. Modified focal point arrangements that decrease the focal point density along the array normal to further reduce normal tissue heating will also be evaluated. Simulations demonstrating the results of these and other modified beamforming strategies will be demonstrated for a large spherical section array using the FOCUS software package.

Session 2pNCa

NOISE-CON and Noise: Construction Noise

Erich Thalheimer, Chair
Parsons Brinckerhoff, 3340 Peachtree Rd., Ste. 2400, Tower Place 100, Atlanta, GA 30326

Contributed Papers

1:00

2pNCa1. Remote construction noise monitoring. Areg Gharabegian (Parsons, 100 W. Walnut St., Pasadena, CA 91124, areg.gharabegian@parsons.com)

Los Angeles was constructing an underground 12 ft in diameter sewage tunnel stretching 25 km. The project includes nine construction access shafts where tunnel boring machines, muck trains, and materials were lowered into the ground. The excavated tunnel material was transported from the tunnel head to the access shafts by muck train, removed by a crane, and carried away via heavy haul trucks. A state of art remote construction noise system was developed and installed near each of the shaft sites, which was collecting noise, audio, and video data 24 h a day. Noise levels were recorded whenever a noise limit threshold is exceeded. Data were then transmitted via modem to a central location where noise specialists review and analyze the data. A portion of these noise limit exceedances were correlated to the audio and video files to determine if the exceedances were due to the construction activities or other non-construction related events. Construction noise at the noise sensitive sites adjacent to each shaft were minimized by constructing sound walls as high as 10 m.

1:15

2pNCa2. Rising to the challenges of noise monitoring around construction sites. Douglas Manvell and Martin Alexander (Bruel & Kjaer, Skodsborgvej 307, 2850 Nrum, Denmark, dmanvell@bksv.com)

The construction of infrastructure and new buildings causes changes to the environment of the neighborhood both during and possibly after construction, particularly for major projects. These changes are often significant and result in concern among local population and government and often result in restrictions. To be effective and thus meet the demands on these restrictions, adequate enforcement is needed. Enforcement of noise limit compliance around construction sites needs to efficiently determine noise limit compliance, avoiding false positive exceedences, in order to provide a correct and true picture of the impact of the construction work. This paper describes some of the major challenges faced when designing, deploying, and operating a system to monitor compliance of noise regulations and limits around a construction site. It describes and compares alternative approaches to facing these challenges.

1:30

2pNCa3. Remote construction vibration monitoring at sensitive facilities. Marc Newmark, Jeffrey Zapfe, and Eric Wood (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138, mnewmark@acentech.com)

Construction-related vibrations, while temporary, still have the potential to disrupt vibration- or noise-sensitive operations in adjacent facilities (such as hospitals, laboratories, and performing arts venues). Traditional construction vibration monitors are configured to measure levels corresponding to building damage thresholds. Unfortunately, vibration-sensitive equipment can be adversely affected at levels that are orders of magnitude lower than this. Furthermore, while most construction vibration monitors provide level information at a single dominant frequency, most criteria for sensitive equipment are specified over a range of frequencies. Generally, occupants of buildings near construction sites prefer to continue operations during the construction process. To do this, however, they need to know how the construction-related vibrations compare to their instrument criteria. To help provide this information, we have developed a remote monitoring system that provides real-time vibration and/or noise spectra in the sensitive facility. These spectra can be directly compared to instrument criteria with appropriate alarm notifications as needed. In this paper, we will present three case studies that show the evolution of this remote monitoring capability and discuss how the information has been used by the institutions where it has been installed.

1:45

2pNCa4. An approach to noise impact assessment for long-term construction. Frank Babic (AECOM, 105 Commerce Valey Rd., West, Markham, ON L3T 7W3 Canada, frank.babic@aecom.com)

Noise generated by construction is typically characterized as a temporary noise source. Due to its temporary nature, construction noise can generally be tolerated by the community at large more readily than long-term noise sources (for example, from industry). For some projects, construction may extend beyond a temporary period to become a more long-term noise problem—especially when the construction activity can last months or even years. When an extended construction period occurs, the community may be less tolerant of the noise, and the assessment of its impact on the community is more aptly characterized as a long-term noise source, similar to that of industry. This paper proposes an approach to assessing long-term construction projects, where a "worst-case" scenario can be modeled with utility factors, duty cycles, and barrier noise controls. From this modeling, a zone of influence can be determined which identifies the extent of the noise impact (i.e., potential complaints generated by nuisance from construction noise). Receptors within this zone of influence (including residences, hospitals, churches, etc.) can then be identified as part of the environmental noise impact assessment for long-term construction activity.

2:00

2pNCa5. Estimating construction noise using the Environmental Science Research Institutes geographic information systems program. Scott Noel (Parsons Brinckerhoff, Inc., 400 SW 6th Ave., Ste. 802, Portland, OR 97219, noels@pbworld.com)

ArcGIS is a series of Geographic Information Systems (GIS) programs and tools developed by the Environmental Science Reaserch Institue (ESRI). ArcGIS has many tools built in, however, no tools to extimate construction noise levels for either point sources or for a project construction footprint. PYTHON is a programming language that is becoming the preferred programming language for ESRI programs. Parsons Brinckerhoff, Inc. has modified the buffer tool in ArcGIS to estimate noise levels from a variety of construction equipment. The tool allows the user to choose what typical construction equipment will be used on their project. The tool then creates a noise contour showing what noise levels could be expected at certain distances from the project area or point source. The tool is useful for estimating construction noise at various distances to sensitive receptors. The tool also helps in graphics creation to aid in summarizing noise impacts. Parsons Brinckerhoff, Inc. has used this tool on transportation projects but it is applicable to other projects as well.

2:15

2pNCa6. Construction noise impact prediction using non-virtual simulation. Rob Greene (Parsons Brinckerhoff, 505 S. Main St., Ste. 900 Orange, CA 92868, greener@pbworld.com)

This paper presents a rarely used approach to predicting potential noise impacts from major infrastructure construction. The method includes the actual propagation of spectrally shaped pink noise into a community from locations within the proposed project's construction site, coupled with contemporaneous noise measurements at multiple in-community locations. Pretest source calibration measurements were performed and ambient noise levels were measured in the surrounding community during pre- and posttest periods. A Public Information program was part of the environmental assessment process. Yes, this novel/brave/are you crazy? approach was extensively covered in the local press. The community reaction to the methodology, the noise control mitigation that resulted from the test, the success of the construction, and lessons learned will be presented.

2:30

2pNCa7. Using the Federal Highway Administration Roadway Construction Noise Model for the Jerome Park reservoir construction project. Gabriella Yanez-Uribe and Erich Thalheimer (Parsons Brinckerhoff, 3340 Peachtree Rd, Ste. 2400, Tower Pl. 100, Atlanta, GA 30326, yanezuribe@pbworld.com)

The new Jerome Park Reservoir in the Bronx, NY, is a critical water supply source for the residents and businesses of New York city. This paper will describe the noise modeling work done in support of the project as well as an odd bug found in the FHWA's RCNM model. Noise associated with the various construction phases and equipment for the project were analyzed to evaluate the potential loudest hour during each month of work for the project's 50-month duration. Noise analyses were performed for expected weekday work conditions and Saturday work conditions at several receptor locations surrounding the reservoir using the FHWA's Roadway Construction Noise Model (RCNM), Version 1.0. During the analysis, an error was found with the output function of RCNM when exporting results for multiple receptors into CSV and TXT output files. The total L_{eq} and L_{max} levels were being reported as zeros even though the correct total noise levels were displayed on screen. This error was submitted to FHWA for review, and as a result, RCNM Version 1.01 was soon released addressing the problem. More importantly, the project was performed to the client's satisfaction and has allowed the necessary construction work to proceed.

2:45

2pNCa8. How good is the Roadway Construction Noise Model for predicting highway construction noise impacts? Marc Wallace (Tech Environ., Inc., 303 Wyman St., Ste. 295, Waltham, MA 02451, mwallace @techenv.com)

The reconstruction of the I-93/Route 110/Route 113 in Methuen, MA will be conducted over a 4-year period. As part of the project, a new I-93 bridge superstructure will need to be built. This bridge work will consist of constructing the center portion of new superstructure, which will include pile driving. The work will require shifting I-93 traffic lanes several times to accommodate bridge construction over a several months. The design engineers are also considering constructing the superstructure over a weekend by shutting down I-93 and re-routing over local roads to reduce construction impacts. The Federal Highway Administration's (FHWA) Roadway Construction Noise Model (RCNM) was used to assess potential construction noise impacts on nearby residential neighborhoods. The RCNM is a good screening tool to conservatively assess potential construction noise impacts. There are more sophisticated models, such as the Cadna A comprehensive three-dimensional (ISO 9613-2) model that takes into account topographic and atmospheric conditions. The goals of this paper will be to compare the RCNM and Cadna A model results to assess the accuracy of the RCNM model and to also assess the potential benefits or impacts of constructing the superstructure over a shorter period of time using the Cadna A model.

3:00

2pNCa9. Proactive regulation engenders creative innovation: Quieting the jack hammer. Eric Zwerling (Dept. of Environ. Sci., Rutgers Noise Tech. Assistance Ctr., 14 College Farm Rd., NB, NJ 08901, Zwerling@rutgers.edu), Charles Shamoon (New York City Dept. of Environ. Protection), and Stephen Szulecki (The Noise Consultancy, LLC)

In 2002, the New York City Department of Environmental Protection undertook to completely revise the city's Noise Code; it went into effect in 2007. The Mayor's Office made the reduction in construction noise a priority. Incentives were provided for the application of best management practices and the use of quieter equipment. DEP staff proposed the development of a remediation strategy for pneumatic pavement breakers. Working with KeySpan Energy (now National Grid), they approached Zo-Air Company, an industrial tool distributor, who developed the "No Racket Jacket." The key to developing a successful engineered solution is understanding the noise source. Most of the noise emissions from a pavement breaker come from internal components, not the tool bit hitting the pavement. Enclosing the source significantly reduces the noise emissions. The jack*et al.*so serves to insulate the operator from the exhaust air of pneumatic units, and the oil often spewed onto their legs. This paper discusses the development process which included four prototypes, and the results, with data from source sound level measurements and operator dosimetry. Three scenarios were tested: unremediated pneumatic hammer, pneumatic hammer with the "No Racket Jacket," and an electric pavement breaker.

3:15—3:30 Break

3:30

2pNCa10. Designing partial enclosures to minimize airborne sound transmission. D. W. Herrin, A.E. Carter (Dept. of Mech. Eng., Univ. of Kentucky, 151 RGAN Bldg., Lexington, KY 40506-0503, dherrin@engr.uky.edu), J., Shi, and D. C. Copley (Caterpillar, Inc.)

Enclosures are a very common way to reduce noise emissions from machinery. However, enclosures display complex acoustic behavior that is difficult to predict. Boundary element simulation was used to better understand the airborne transmission path for a partial enclosure. Sources considered included a point source and a diesel engine. Insertion loss was used as the performance measure to evaluate the effect of several different design considerations. Results indicate that the most important factors affecting enclosure performance are the opening size, amount of absorption, and the source-to-opening distance. It is notable that insertion loss was relatively insensitive to enclosure or source size.

3:45

2pNCa11. New York City western rail yard construction noise study. Erich Thalheimer (Parsons Brinckerhoff, 75 Arlington St., Boston, MA 02111, thalheimer@pbworld.com)

This paper will describe the technical approach, regulatory environment, and outcome of a construction noise study performed on behalf of a $2.5 billion private development to be built over the Western Rail Yards in Manhattan. Community noise modeling for the 9 year project was performed in accordance with New York City Environmental Quality Review (CEQR) Manual requirements using the CADNA-A and FHWA TNM models. The on-site noise analysis accounted for the staggered schedule of excavation, foundation laying, erection, and finishing of eight high rise residential and commercial buildings. Construction equipment source emissions used in the CADNA-A model were taken from the FHWA Roadway Construction Noise Model (RCNM) and the NYC DEP Construction Noise Regulations (Local Law 113, Chap. 28). Project-related trucking (mobile) noise was evaluated using the TNM model. Seventeen multi-story noise receptor locations surrounding the development site were evaluated with ambient noise measurements and predictive noise modeling. A construction vibration assessment for the elevated High Line park was also conducted. The results indicated that noise exceedance conditions were expected from pile driving, hoe ramming, and jackhammering operations during daytime, nighttime, and weekend periods, so noise mitigation measures were developed and incorporated into the project's construction permit.

4:00

2pNCa12. Report on mitigation of nighttime construction noise. Cliff Schexnayder (Arizona State Univ., P.O. Box 6700, Chandler, AZ 85246, cliff.s@asu.edu) and Jennifer S. Shane (Iowa State Univ.)

Nationally, the amount of road work being done at night appears to be significant. Data taken during the summer of 2001 suggests that one-third of the work zones present on National Highway Systems roadways were active primarily at night. Night work can have impacts on adjacent neighborhoods and businesses that would not be experienced or as disturbing if conducted during the daytime. Construction equipment is a major noise and nuisance generator on nearly all nighttime construction projects. This paper will report on an ongoing Transportation Research Board study that is seeking to identify best practices in terms of source control, path control, and receptor control of noise on transportation projects. Individual DOTs having specific experience and contractors are being queried about their experiences in mitigating nighttime construction nuisances.

4:15

2pNCa13. Noise control evaluation for a concrete batch plant. Robert D. O'Neal (Epsilon Assoc., Inc., 3 Clock Tower Pl., Ste. 250, Maynard, MA 01754, roneal@epsilonassociates.com)

As a result of neighborhood complaints, an Administrative Consent Order was issued to a concrete batch plant for violating the Massachusetts noise policy. The fact that the batch plant had been in business long before the residential subdivision was built did not factor in the State's decision. A detailed noise control study was conducted for the facility. All mobile and stationary sources of noise were measured in addition to measurements at the receiving property. The major sources included mixer trucks receiving concrete and prepping their load, concrete trucks climbing a hill, batch plant dust collector fans, and bulk cement truck deliveries. The measurements were analyzed, and a series of noise control measures was evaluated and recommended, some of which were implemented by the facility. Several rounds of controls and associated sound level testing were conducted to determine if the batch plant met the State noise requirements, and compliance was eventually achieved. This paper will discuss the key sources of noise at the facility and the techniques used to reduce noise at the source.

4:30

2pNCa14. Road construction noise inside a radio station. Sean Connolly (Big Sky Acoust., LLC, P.O. Box 27, Helena, MT 59624, sean @bigskyacoustics.com)

Construction noise limits are typically presented as noise levels not to be exceeded at the exterior of a building or a property line. However, when a building houses noise-sensitive uses, the possible construction noise impacts must be assessed inside the building. In 2004, a contentious legal dispute erupted between a radio station owner and the state highway department over the reconstruction of the highway directly adjacent to the station. The owner claimed that excessive road construction noise would disrupt live broadcasts from his studio. Therefore, the construction noise levels needed to be assessed inside the studio during the entire construction process, but legal maneuvering severely limited access to the station. This paper discusses the criteria and methods used to assess construction noise impacts inside the studio, the relationship between the measured noise levels inside and outside of the studio during various phases of road construction, and of course, if the construction noise interfered with radio broadcasts.

TUESDAY AFTERNOON, 20 APRIL 2010 GRAND BALLROOM III/IV, 1:00 TO 3:30 P.M.

Session 2pNCb

NOISE-CON and Noise: Information Technology Noise

David Nelson, Cochair
Nelson Acoustics, P.O. Box 879, Elgin, TX 78621

Terry W. Baird, Cochair
Hewlett-Packard Company, 11311 Chinden Blvd., Boise, ID 83714

Marco Beltman, Cochair
Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124

Contributed Papers

1:00

2pNCb1. German information technology-equipment noise categorization concept: Study on several data projectors. Gregor Feneberg (Mueller-BBM GmbH, Robert-Koch-Str. 11, 82152 Planegg, Germany, gregor.feneberg@muellerbbm.de)

An investigation project tendered by the German Federal Institute for Occupational Safety and Health (BAuA), with a planned duration of 1 1/2 years, has recently been awarded to Mueller-BBM. The existing concept of the quality classification of household appliances, where the energy efficiency categories A–F each represents a determined range of power consumption, is to be transferred to a system for the categorization of IT-equipment with regard to its noise emission. Currently it is almost impossible for a layperson to compare noise declarations provided by manufacturers even though there are standardized and very detailed specifications in ISO 9296 for the measurement and declaration of the noise emission of IT-equipment. One of the central issues of the project is the consolidation of existing procedures for noise emission measurements on IT-equipment with consideration of the impact of psychoacoustic parameters. A categorization proposal based on ISO 7779 will be presented. In order to check the neccessity of additional psychoacoustic metrics like tonality and transient loudness, 11 data projectors were measured according to the BAuA proposal. In addition, the annoyance of the operational noise was evaluated by a panel test. The results will be compared and contrasted with the current ISO /ECMA standard measurements.

1:15

2pNCb2. Development of the standard ISO 7779 3rd edition: Revised and new contents. Ikuo Kimizuka (IBM Japan, Ltd., 1623-14, Shimotsuruma,Yamato-shi, Kanagawa-ken 242-8502, Japan, kmzk@jp.ibm .com)

ISO 7779 is a noise test code for airborne noise emission of information technology and telecommunications equipment (so called, ITTE). Since its 1st edition of 1988, ISO 7779 has stated, in detail, the installation and op-

erating conditions for specific equipment categories. In nature of ITTE, however, changing of the product characteristics is much faster than the pace of standard document updates. For instance, most models of copy machines and printers (except some dedicated high speed machines) are almost being combined into new category, so called, multi-functional devices. To catch such rapid changes up timely, it becomes impractical to state specific condition within this International standard. Therefore, ISO/TC43/SC1 /WG23 which is responsible for ISO 7779 decided to propose new frame work so that 3rd edition should refer to ECMA-74, annex C, which has been the industrial counterpart and also pilot standard of ISO 7779. This paper introduces background of key changes from 2nd edition and new technical contents for ITTE noise emission measurement methodologies.

1:30

2pNCb3. Printer impulsive noise metrics: Sensitivity and repeatability analysis. Terry Baird and Katy Ferguson (Hewlett-Packard Comp., 11311 Chinden Blvd., Boise, ID 83714, terry.baird@hp.com)

The major cause of annoyance from impulsive noise is the impulsive amplitude. Many methods can characterize impulsive amplitude, such as time-averaged measures, peak amplitudes or real-time levels. For example, the "impulsive noise index" described by ISO 11201 or the "impulsive parameter" described in ISO 7779 are time-averaged metrics. Real-time metrics might include peak sound pressure level or peak loudness in the time domain. Impulsive metrics must be evaluated for effectiveness. That is, the metric must agree with the perceived impulsive amplitude. Impulsive metrics must also be practical. They must be insensitive to sample-to-sample variation while still providing the granularity between passing and failing results. This paper focuses on evaluating the sensitivity, measurement repeatability, and result granularity to determine the practicality of various effective impulsive noise metrics.

1:45

2pNCb4. Development of schemes to gauge the annoyance of time-varying sound. Menachem Rafaelof (Seagate Technol., 389 Disc Dr. Longmont, CO 80503, menachem.rafaelof@seagate.com)

Hard disk drives (HDDs) nowadays have use in many diverse applications. Examples include various types of consumer electronic devices that operate in extremely quiet environments and personal computers that integrate the HDD and other peripherals within the display very close to the user. Higher exposure to the HDD sound, either due to the lack of masking sound or close proximity to the user, require establishment of a scheme to identify features within the sound to decide on its acceptability. This paper describes an effort aimed at developing a scheme for identification of attention-grabbing or annoying features within time-varying sound that would point to its unacceptability by a typical user. In this first part, sample sounds were examined based on their perceived strength (loudness), duration, repetition rate, and higher order statistics to acceptability based on a jury study.

2:00

2pNCb5. Impact of personal video recorder/digital video recorder on hard disk drive acoustic requirements. Dave Ali (Western Digital, 5863 Rue Ferrari, San Jose, CA 95138, dave.ali@wdc.com)

In recent years the impact of the personal video recorder/digital video recorder market has grown substantially. This growth has a direct impact on the acoustic test laboratories that make sure the systems meet the manufacturer's specification for noise emissions. To date the standards have not addressed this growth and subsequently have not provided guidance as to standardized methods of test for these systems. This paper will show the impact of the market and provide a proposed addition to the appropriate acoustic test standards.

2:15

2pNCb6. Fan noise prediction for design. Charles Oppenheimer (Hewlett-Packard, 18110 SE 34th St., Vancouver, WA 98683, charles .oppenheimer@hp.com)

This paper develops a fan noise prediction model intended for use in the design of systems for cooling and drying. Airflow through a system is described by a dimensionless resistance parameter. Fan noise is predicted by a scaling relationship obtained from a dimensional solution of Lighthill's acoustic analogy. The scaling relationship contains unknown proportionality factors, which are determined empirically and compared for axial fans and centrifugal blowers. The dependence of fan and blower noise spectra on non-dimensional scaling parameters is also investigated.

2:30

2pNCb7. Constant acoustic fan curves for 92 mm axial fans. Willem Beltman (Intel Labs, Intel Corp., M/S JF2-86 2111 NE, 25th Ave., Hillsboro, OR 97124, willem.m.beltman@intel.com)

The noise emission of fans changes as a function of the operating point. Airflow and acoustic fan plenum experiments can be used to characterize these effects to estimate noise under installed conditions. Recently, the concept of iso-acoustic fan curves was introduced by various authors, and results were presented for small radial blowers and some axial fans. The current paper presents an analysis of the iso-acoustic fan curves for an extensive set of 92 mm axial fans. An analysis is presented of the airflow and acoustic performance of these fans, and expressions for the iso-acoustic fan curves are derived and presented. This allows a direct comparison between the various fans based on operating point conditions.

2:45

2pNCb8. Aero-acoustic simulations of small scale electronic cooling blowers. Jessica Gullbrand and Willem M Beltman (Intel Labs, Intel Corp., 2111 NE 25th Ave., MS JF2-86, Hillsboro, OR 97124, jessica.gullbrand @intel.com)

The noise emission of small radial blowers is important for the cooling of notebook systems. It has been demonstrated that airflow simulations for these radial blowers can be performed to predict the blower performance, even in the presence of inlet restrictions. The current paper presents an extension of this work to include predictions of the radiated acoustic noise. The calculated blower flow field is used as an input for a second stage aeroacoustic prediction. This paper outlines two methods. The first method uses a free field approach based on the Ffowcs-Williams and Hawkings analogy. The second approach constructs a rotating fan source, the magnitude of which is determined by integrating the pressure across a number of blade segments. The radiated field is then determined in a boundary element calculation, both for a free field and in the presence of the fan enclosure. Results for both methods are presented and compared to experimental data.

3:00

2pNCb9. A relation between flow, noise, and system impedance for notebook computers. Eric Baugh (Intel Corp., 15400 NW Greenbrier Pkwy., M/S CO5-166 Beaverton, OR 97006, eric.baugh@intel.com)

Performance of notebook computers is strongly limited by the ability to reject heat from the system. In a typical mainstream design, the airflow driven by the internal fan is responsible for about 80% of the maximum cooling capability, and this airflow impacts all the internal components (CPU, memory, radios, etc.) as well as the top and bottom skin temperatures. Fan flow, in turn, is limited by acoustics. It has been shown that the fan installation conditions inside a notebook computer have an adverse effect on both flow rate and noise. The market trend is toward thinner systems, which generally have a higher system impedance. As impedance increases, system airflow must decrease to maintain a fixed acoustic limit, holding all other variables constant. In order to demonstrate the effect of impedance, this paper describes a simple, two parameter model relating the flow rate in a notebook computer at a given operator position sound pressure level to the impedance index of the system. This relation is shown to reasonably approximate the behavior of about 20 different notebook computers.

3:15

2pNCb10. The dB[EQL]: An alternative sound pressure weighting according to the equal loudness contours of the international standard ISO 226-2003. Wade Bray (HEAD Acoust., Inc., 6964 Kensington Rd., Brighton, MI 48116, wbray@headacoustics.com)

The conventional spectral weightings applicable to sound pressure [dB (A), etc.] are fixed in spectral shape and intended for use over certain ranges

of unweighted sound pressure level: for example, the *A*-weighting from threshold of hearing up to 65 dB(SPL). In general, these weightings' spectral shapes match the sensitivity of hearing as a function of frequency within "use-level" ranges, although the strong effect of the cavum conchae resonance of the ear, apparent in the equal-loudness contours of ISO 226, is not considered except in the rarely Used *D*-weighting. Particularly for sounds with tonal content and within the general level range of the *A*-weighting, the authors propose a new spectral weighting assembled from the Phon values of the complete set of equal-loudness contours calculated for each frequency within the human auditory range. It will be shown that although giving values similar to those of the *A*-weighting, the dB[EQL] or equal-loudness weighting is situation dependent rather than fixed, and better represents subjective impressions at all frequencies. Although based on perceived loudnesses, the dB[EQL] sound pressure weighting is not a specific loudness measurement, does not consider critical band formation, and does not yield masking or psychoacoustic loudness data.

TUESDAY AFTERNOON, 20 APRIL 2010 GRAND BALLROOM III/IV, 3:45 TO 5:00 P.M.

Session 2pNCc

NOISE-CON and Noise: Consumer and Industrial Product Noise

Charles S. Hayden, Chair
NIOSH, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998

Contributed Papers

3:45

2pNCc1. Geothermal heat pump noise control (cont.): Its all in the details. Jeffrey Fullerton (Acentech, Inc., 33 Moulton St. Cambridge, MA 02138, jfullerton@acentech.com)

The installation of a geothermal heat pump in the basement of a 1900 era home introduced significant noise and vibration to a bedroom above, which was discussed in detail in a previous conference. After mitigating the noise and vibration using piping isolation products, lower levels of noise and vibration still remained. This paper will discuss pursuits that were implemented and the very effective resolution for mitigating the noise and vibration that was generated by the heat pump.

4:00

2pNCc2. Improving screw compressor housing design using simulation. D. W. Herrin, Z. Tao (Dept. of Mech. Eng., Univ. of Kentucky, 151 RGAN Bldg., Lexington, KY 40506-0503, dherrin@engr.uky.edu), A. J. Graybill, J.E. Bender, P.C. Marks, and C. Eichelberger (JCI Bldg. Efficiency)

The sound radiation from a rotary screw compressor was simulated using structural finite element and acoustic boundary element analysis. The modes of the housing were measured with and without the screws. Finite element modal analysis results were compared to the experimental modes with good agreement. Results demonstrated that the screws affected the bending modes but not the more important cylinder modes of the compressor housing. The sound radiation from the compressor housing was then predicted using acoustic boundary element analysis. The results indicated that the second and fourth pumping frequencies were close to two structural modes. Subsequent simulations investigated the effect of design changes on the housing. These changes included using a thicker cylinder and adding ribs.

4:15

2pNCc3. Vibro-acoustic model of an enclosed electric power generator. David C. Copley and Harvind Raman (Caterpillar Inc., Tech. Ctr. Bldg. E, P.O. Box 1875, Peoria, IL 61656, copley_david_c@cat.com)

A vibro-acoustic model of an enclosed diesel-powered electric generator was built using commercial vibro-acoustics software (primarily SEA) and empirical methods. The vibro-acoustic sources, including the engine, generator, cooling fan, and exhaust were characterized experimentally for several different operating conditions. The vibro-acoustic performance of the enclosure and sound suppression features were modeled and experimentally validated. Good agreement was achieved between experimental and simulated results.

4:30

2pNCc4. Passive noise control methods applied to a household icemaker. Richard Ruhala (Southern Polytechnic State Univ., 5735 Forest Ave., Evansville, IN 47712, rruhala@usi.edu)

Two icemakers were provided to the engineering program at the University of Southern Indiana for educational testing and development. Three students worked sequentially over 1 year to quantify and localize the noise sources using sound pressure and sound intensity methods, and then implement various design modifications in order to reduce the overall sound power by a goal of 3 dBA. Noise reduction was achieved during some machine operating cycles, although the overall noise reduction goal was not obtained. The students did learn one important finding in noise control engineering: that reducing the noise of a well developed and quiet product is a difficult task that requires expertise in noise assessment and control. This paper will also focus on findings from these studies that are pertinent to the development of quieter refrigerators, freezers, and icemakers.

4:45

2pNCc5. Acoustics of energy efficient hand dryers: Is this progress? Jeffrey Fullerton and Gladys Unger (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138, jfullerton@acentech.com)

Recently there have been innovations in the technology of bathroom hand dryers with a particular focus on energy efficiency. The results have included some rather interesting methods for drying wet hands in different ways than previous generations of hand dryer models. The focus of this study will be to compare the sound levels from these different energy efficient models with previous models that have been used for years. The overall sound level will be compared to the rated efficiency of each unit.

Session 2pNCd

NOISE-CON, Noise, and Physical Acoustics: Aircraft Exterior Noise

Eric R. Boeker, Chair
USDOT Volpe NTSC, 55 Broadway, Cambridge, MA 02142

Contributed Papers

1:00

2pNCd1. FAA noise impacts research roadmap. Raquel Girvin (FAA, 800 Independence Ave., SW Washington, DC 20591, raquel.girvin@faa.gov)

The Federal Aviation Administration's (FAA) continuing mission is to provide the safest, most efficient aerospace system in the world. Improving efficiency through airspace redesign, airport capacity expansion, and other initiatives of the FAA Next Generation Air Transportation System (NextGen) may be hampered without an aggressive program to address the environmental consequences of aviation. FAA's work toward mitigating such environmental consequences includes research to advance the state of knowledge on noise impact analysis. This paper describes the status of FAA's ongoing research efforts and the comprehensive research roadmap addressing critical noise impacts research needs.

1:15

2pNCd2. Overview of aircraft en-route noise prediction. Bill (Hua) He (Federal Aviation Administration, Office of Environment and Energy (AEE) —Noise Div., 800 Independence Ave., NW, Bldg. 10A, 900West Washington, DC 20591, hua.he@faa.gov) and Eric Boeker (US Dept. of Transportation Volpe Ctr.)

Community noise near airports is a major aviation noise concern, prompting the development of various modeling tools to predict the noise changes due to airport expansion and fleet changes. Once aircraft leave the terminal area and reach high altitudes, the aircraft noise received on the ground is assumed to be insignificant relative to urban ambient noise and is often neglected. However, this might not be the case, under either of the following conditions: (1) when aircraft fly over areas with very low ambient noise levels, such as in wildness areas, or (2) when an aircraft is powered by advanced propulsion systems that produces high noise levels. There is an increasing need to address both of these issues. Over the past 2 decades, some limited noise modeling capabilities and related research have begun to quantify the high-altitude aircraft noise or "en-route" noise. The purpose of this paper is to review and summarize related noise prediction research conducted through direct technical exchange with researchers, some simple data analysis as well as through a literature review. The authors will draw on experience in developing current aircraft community noise tools and formulate key future research questions and options.

1:30

2pNCd3. Application of atmospheric absorption models for aircraft enroute noise. Kieran Poulain, Victor Sparrow, Kenneth Brentner (201 Appl. Sci. Bldg., Univ. Park, PA 16802, kxp252@psu.edu), and Louis Sutherland (Los Angeles, CA)

Propagation of aircraft enroute noise for flights above 18 000 ft above ground level (AGL) (5.49 km AGL) is important for estimating noise impact in U.S. National Parks and other quiet areas. One key to the accurate estimation of noise impact is atmospheric absorption. In our atmosphere, it turns out that the absorption coefficient is dependent on altitude as well as frequency. The altitude dependent parameters include mean pressure, temperature, and relative humidity as well as the concentration of molecular species. This altitude effect can be important for certain frequency bands, and this point will be emphasized in this paper. In addition, an improved atmospheric absorption model for high altitudes of Sutherland and Bass [J. Acoust. Soc. Am. **115**, 1012–1032 (2004)] shows slightly different values of atmospheric absorption in important frequency bands for aviation noise at altitude compared to the typical ANSI Standard S1.26-1995 (R 2009). For certain situations, the updated atmospheric absorption coefficient could produce different enroute noise predictions on the ground. [Work supported by FAA/NASA/Transport-Canada PARTNER Center of Excellence.]

1:45

2pNCd4. Modeling propagation of aviation noise under range-dependent conditions, using a hybrid parabolic equation-fast field program method. Joyce Rosenbaum, Anthony Atchley, and Victor Sparrow (Graduate Program in Acoust., The Penn State Univ., 202 Appl. Sci. Bldg., University Park, PA 16802, jer262@psu.edu)

Based on the Parabolic equation method, a model for sound propagation over uneven terrain, with mixed ground impedance and range-dependent meteorological conditions, has been developed to improve prediction of the impact of aviation noise on communities. Intended to enhance the noise prediction capabilities for the Federal Aviation Administration's Aviation Environmental Design Tool in support of the Next Generation Air Transportation System (NextGen), the model uses a hybrid parabolic equation-fast field program approach for computing aviation noise contours. These methods ensure that the low-frequency content, a factor in community impact, is propagated accurately. The model can accommodate a broadband, moving sound source, traveling along a user-specified path, over three-dimensional terrain. A test case, designed to represent simple but realistic propagation conditions, is described, and results for a source traveling along a simple flight path are presented and discussed.

2:00

2pNCd5. Acoustic repropagation technique and practical source characterization for simulation noise model databases. Christopher Hobbs and Juliet Page (Wyle Labs., 241 18th St. South, Ste. 701, Arlington, VA 22202, chris.hobbs@wyle.com)

By utilizing the acoustic repropagation technique in conjunction with an open-source, platform-independent database, a robust noise source can be constructed that is well suited for use in simulation noise models. This paper describes a widely applicable methodology for creating a three dimensional noise source from measurements that can account for source directivity. It will be shown how the resulting data, known as noise spheres, may contain a variety of noise metrics representing the source at a fixed distance for all spherical angles. Such spheres can be combined to represent complex noise sources and depending on their metric content can be used to accurately predict the near and far field environment about the source. A case utilizing narrow band spectra with phase information will be explored. It will be shown how a collection of spheres can be used to represent different operating states of a source and how the use of an independent parameter such as speed or power can be used to interpolate non-exact states in a noise model. By way of example, an approach to incorporating non-linear sound propagation from a high amplitude noise source is explored in conjunction with the advanced acoustic model.

2:15

2pNCd6. Modeling of noise from aircraft taxiing operations. Juliet Page and Christopher Hobbs (Wyle Labs., 241 Eighteenth St. South, Ste. #701, Arlington, VA 22202, juliet.page@wyle.com)

This paper documents a study performed for the Airport Cooperative Research Program, investigating noise from aircraft taxi operations. A holistic approach is utilized, applying an understanding of the operations, the detailed aircraft motions, and the acoustic sources. A series of acoustic sensi-

tivity studies was conducted in order to develop a physical understanding of the relative importance of the various modeling elements within the framework of AEDT. The study is based on decoupling the modeling of taxi noise into the following three areas and exercising each element independently. (1) Engine source noise (level, spectra, and directivity); (2) aircraft movements and operating states (location, duration, and power setting); (3) environment/propagation (lateral attenuation, terrain, shielding, and ground impedance). Aircraft taxi noise sensitivity studies exercised each element independently. Limited opportunistic commercial aircraft taxi operation acoustic measurements were conducted. Independent taxi flight data recorder information was queried to determine statistical engine and aircraft operational parameters. The sensitivity studies reveal that the primary weakness for taxi noise modeling is related to item (#1), engine source noise modeling. This paper will discuss the sensitivity study findings and present an approach for development of a nominal taxi state noise noise-power-distance, spectral class, and directivity database for AEDT.

2:30

2pNCd7. Investigation of the use of lower frequency acoustic data in helicopter environmental noise modeling. Eric Boeker and Noah Schulz (USDOT Volpe Natl. Transportation Systems Ctr., 55 Broadway RVT-41 Cambridge, MA 02142, eric.boeker@dot.gov)

The current Federal Aviation Administration (FAA) aircraft noise modeling tools, Aviation Environmental Design Tool (AEDT) and Integrated Noise Model (INM), do not consider noise below 50 Hz in their computations. This paper describes a preliminary study to determine the effect of including low-frequency data on the accuracy of AEDT/INM results. Expanded aircraft noise spectra containing one-third octave band data to 12.5 Hz were analyzed using methods adapted from AEDT/INM. Results from expanded spectral data are compared with results from the historical AEDT/INM spectral data (one-third octave band data from 50 Hz to 10 kHz). This comparison showed a range of differences, from increases in overall un-weighted sound pressure levels, to negligible changes in *A*-weighted and time audible metrics. These changes may be particularly important for helicopters, with dominant low-frequency rotor noise below 50 Hz. Following the comparison, the potential implementation of expanded spectral data in AEDT/INM is discussed.

2:45

2pNCd8. Examination of the low frequency limit for helicopter noise data in aviation environmental design tool and integrated noise model. Eric Boeker and Noah Schulz (USDOT Volpe Natl. Transportation Systems Ctr., 55 Broadway RVT-41, Cambridge, MA 02142, eric.Boeker@dot.gov)

The current Federal Aviation Administration (FAA) aircraft noise modeling tools, Aviation Environmental Design Tool (AEDT) and Integrated Noise Model (INM), do not consider noise below 50 Hz in their computations. This paper describes a preliminary study to determine the effect of including low-frequency data on the accuracy of AEDT/INM results. Expanded aircraft noise spectra containing one-third octave band data to 12.5 Hz were analyzed using methods adapted from AEDT/INM. Results from expanded spectral data are compared with results from the historical AEDT/INM spectral data (one-third octave band data from 50 Hz to 10 kHz). This comparison showed a range of differences, from increases in overall un-weighted sound pressure levels, to negligible changes in *A*-weighted and time audible metrics. These changes may be particularly important for helicopters, with dominant low-frequency rotor noise below 50 Hz. Following the comparison, the potential implementation of expanded spectral data in AEDT/INM is discussed.

TUESDAY AFTERNOON, 20 APRIL 2010 GRAND BALLROOM VII/VIII, 3:15 TO 4:45 P.M.

Session 2pNCe

NOISE-CON, Noise, and Structural Acoustics and Vibration: Vibrations and Structureborne Noise in Buildings

Jeffrey P. Kwolkoski, Chair
D. L. Adams Assoc., Inc., 1701 Boulder St., Denver, CO 80211

Contributed Papers

3:15

2pNCe1. Flanking noise transmission in condominiums via steel roof deck. Jim X. Borzym (Borzym Acoust. Consulting & Eng., LLC, 2221 Columbine Ave., Boulder, CO 80302, acoustics@columbine.net)

Noise was found to transmit horizontally via a structure borne path. A lightweight roof consisting of ribbed steel deck spanned two adjacent loft-style condominium residential spaces. This noise transmission path significantly flanked the demising partition wall with known high-transmission loss. Two remedies were attempted to reduce transmission, by adding concrete on top of the roof deck. One remedy was successful.

3:30

2pNCe2. Flanking paths associated with exposed roof decks in condominiums. Jeff Kwolkoski (D. L. Adams Assoc., Inc., 1701 Boulder St., Denver, CO 80211, jkwolkoski@dlaa.com)

An award winning residential/commercial condominium project featured two-story loft condominiums with an industrial look. The building was constructed with steel structure, exposed concrete on metal deck floors, and an exposed metal deck roof. Condominiums were separated by gypsum board on metal stud walls. A number of owners in the building complained of poor sound isolation from adjacent units. Field tests showed that sound isolation of the walls fell below expectations for the design. Subsequent investigation showed that a variety of structureborne and airborne flanking paths contributed to poor sound isolation. Improving sound isolation required a combination of conventional and creative methods.

3:45

2pNCe3. Concrete pier foundation and building design for electron microscope installations with vibration disturbances from train, truck, and bus traffic. Chad Himmel (1705 W. Koenig Ln., Austin, TX 78756, himmel@jeacoustics.com)

A university microscopy and imaging center housing nine instruments including scanning electron microscopes and transmission electron microscopes was proposed for a site near an active railroad and roadways. Low-frequency sound and vibration from trains, trucks, and buses can affect the performance and use of electron microscopes, by disturbing the stability and resolution of the specimen's image. On-site observations and spectrum measurements were conducted to determine the ambient and transient ground borne vibration conditions. Measurements showed that disturbances from trains as well as bus and truck traffic on campus roads significantly exceeded allowable floor vibration tolerances for many of the instruments. Soils exhibited response to disturbances at dominant frequencies around 5 and 10 Hz. This paper addresses the potential for concrete pier-supported floor vibration and structural resonance contributions to ground borne vibration levels and solutions proposed to mitigate the disturbances for sensitive

instrument installations. Graphic diagrams of structural design concepts and results from on-site measurements prior to construction will be discussed. The designs and effectiveness of the microscope rooms, structural treatments, and actively damped, pneumatic vibration isolation systems are discussed, along with post-construction vibration measurement results and photographs.

4:00

2pNCe4. Tri-axial measurement of roadway vibration in multiple research buildings located throughout an urban college campus. Scott Harvey and Josh Curley (Phoenix Noise & Vib., 5216 Chairmans Ct., Ste. 107, Frederick, Maryland 21701, sharvey@phoenixnv.com)

Plans are currently underway to run a mass transit rail line through the center of a major university. Concerns over this significant change to the school's infrastructure were raised by members of the university's faculty, particularly professors questioning whether their research projects and vibration sensitive equipment will be disturbed from vibration generated by all the new train passbys. A study was arranged to investigate the campus existing vibration levels and analyze the potential for a substantial increase in vibration once the rail line is constructed. This was primarily a measure of vibration from traffic sources, especially university buses which operate frequently throughout the campus during all hours. The project involved recording simultaneous tri-axial vibration measurements in numerous locations throughout the campus. This required designing methods to manage equipment, make measurements, organize data, and present results in an efficient, timely, and coherent manner. The data indicate the level of ambient ground and structural vibration already present from daily university activities while serving as a basis for comparison to the increased vibration levels, if any, the newly constructed train will cause.

4:15

2pNCe5. A method for determining the force spectrum produced by mechanical equipment. Andrew Gorton and Chris Papadimos (Papadimos Group, 818 Fifth Ave., Ste. 207, San Rafael, CA 94901, andrew@papadimosgroup.com)

Currently there is no standardized method for determining the force spectrum produced by a piece of mechanical equipment. This information is useful for conducting detailed vibration analysis in buildings, specifying proper vibration isolation for mechanical equipment, and comparing mechanical equipment vibration performance. This paper presents a method for estimating the force spectrum for a piece of mechanical equipment supported on spring isolators. The force spectrum is calculated from the measured vibration response of the sprung mass and the frequency response function of the system. For validation, this procedure was used to determine the force spectrum of a 21 500 CFM air handling unit externally mounted on nominally 1-in. static deflection spring isolators. The calculated force spectrum was then used as a force input to a finite element model of a mass-spring-damper system with the same properties as the sprung air handler. The predicted results compare favorably with the measured vibration response and tend to support that this is a viable method for characterizing the force produced by mechanical equipment.

4:30

2pNCe6. Shipboard structureborne noise by resonant interactions of two torsional vibration modes. Darrell Milburn (J&A Enterprises Inc., 16 Broadway, Salem, MA 01970, milburn@jandaenterprises.com)

In recent years several occurrences of a similar engine-firing rate structureborne noise resonance were observed in the passenger areas of different ships. All were 16V four stroke high speed diesel propulsion systems, including engine vibration isolation mountings and remote reduction gears. The engine firing rate noise path came through the reduction gear mounting feet. Interactions of two torsional vibration modes with nearly equal natural frequencies caused the noise: one mode with the maximum amplitude in the engine crank shaft, and the other in the propeller shaft/reduction gear. Over a narrow range of engine speed, the firing rate energy excited both modes. The resonant torsional vibrations were transmitted to the structure through the reduction gear bearing reactions. In the observed cases, the amplitudes of the torsional stresses at resonance were well within the safe limits of the driveline components, raising no concern to vibration analysts assessing only the risk for mechanical failure. This paper outlines the system parameters related to the occurrences of these resonances, permitting noise control engineers to diagnose this problem or to identify "at risk" installations during the design phase. Noise control recommendations are presented as case studies with examples of the achieved noise reduction results.

TUESDAY AFTERNOON, 20 APRIL 2010 DOVER C, 1:20 TO 4:45 P.M.

Session 2pNSa

Noise, Architectural Acoustics and INCE: Ventilation, Fan, and Duct Noise Control II

Lixi Huang, Cochair
The Univ. of Hong Kong, Dept. of Mechanical Eng., Pokfulam Rd., Hong Kong, P.R. China

Kirill V. Horoshenkov, Cochair
Univ. of Bradford, School of Eng., Great Horton Rd., Bradford, BD7 1DP, UK

Jian Kang, Cochair
Univ. of Sheffield, School of Architecture, Western Bank, Sheffield, S10 2TN, UK

Invited Papers

1:20

2pNSa1. Reduction of blade passing noise of a vertical take-off and landing (VTOL) aircraft. Jie Pan (School of Mehanical Eng., The Univ. of Western Australia, 35 Stirling Hwy., Crawley, WA, pan@mech.uwa.edu.au), Hongmei Sun (The Univ. of Western Australia, Crawley, WA, Australia, hongmei@mech.uwa.edu.au), B. S. Walsh, K. D. Do, P. ONeill, and J. Ranasinghe (The Univ. of Western Australia, Australia)

Rotor-stator interaction has been identified as the dominant noise source of a vertical take-off and landing (VTOL) aircraft developed by Entecho. This paper reports field measurement results of blade passing noise of the VTOL air-craft together with its analysis and control. The blade passing event was simulated in a wind tunnel experiment. The flow speed, rotor position, and rotor-stator spacing were varied with the chord-wise pressure distribution of the lead-ing surface of the rotor blade being measured by an array of six flush

mounted microphones. Results show useful features of the pressure distribution on the rotor blade influenced by an upstream stator, which are used for the analysis and prediction of the sound radiation from the VTOL aircraft. The preliminary result of the reduction in the blade passing noise using angled stator blades is also presented.

1:40

2pNSa2. Prediction of axial fan noise by linearized perturbed compressible equations with large eddy simulation. Young J. Moon, Hoyeon Kim, Youngmin Bae (Dept. of Mech. Eng., Korea Univ., Seoul 136-701, Republic of Korea, yjmoon@korea.ac.kr), Florent Ravelet, Sofiane Khelladi, Hussain Nouri, and Farid Bakir (ParisTech, Paris, France)

The seek for an efficient aerodynamic and aeroacoustic design of axial-flow fans is an important field of investigation for both academic and applied research. Improvements can only be made with a better understanding of the physical mechanisms arising in these machines that combine strong interactions between rotating and non-rotating parts of highly complex geometries. One way is to couple well-suited experimental investigations and innovative computational methods that overtake the weaknesses of methods based, for instance, on aeroacoustic analogy. In this paper we study an axial fan using a new numerical method based on large eddy simulation /linearized perturbed compressible equations with Brinkman penalization. This method is developed in the Department of Mechanical Engineering at Korea University. The experimental tests and validations are performed in the Laboratory of Fluid Dynamics at Arts & Métiers ParisTech in Paris. Detailed analysis of numerical and experimental results is in progress within the two partner teams.

2:00

2pNSa3. The near-field of spinning sources: Why source identification is hard. M. J. Carley (Dept. of Mech. Eng., Univ. of Bath, Bath BA2 7AY, England, m.j.carley@bath.ac.uk)

An asymptotic analysis is presented for the near field of spinning sources, based on a transformation into cylindrical coordinates centered on a line at a fixed radius from the source axis. This transforms the circular source into an equivalent finite length line source with a source distribution made up of "modes" given as Chebyshev polynomials of the second kind $U_n(s) = \sin[(n + 1)\cos^{-1} s]/\sin\cos^{-1} s$. These modes play a role like that of modes in ducts and the analysis shows that their acoustic field propagates or decays depending on whether $n + 1 < k$ or $n + 1 \geq k$, respectively, with k wavenumber, similar to the cut-on/cut-off behavior of duct modes. The analysis is used to examine the problem of identifying a source from field measurements. This has a wide range of applications in many industrial fields but is well known to be (very) ill conditioned. Using the information supplied by the analysis of the source near field, the reasons for this ill conditioning are explained and some possible methods to mitigate the problem are outlined.

2:20

2pNSa4. Toward the prediction of fan noise: From low-speed to high-speed turbomachineries. Stphane Moreau (Dept. of Mech. Eng., Univ. de Sherbrooke, 2500 Boulevard de l'Université, Sherbrooke, QC J1K2R1, Canada stephane.smoreau@gmail.com)

Environmental concern and comfort trigger more and more acoustic specifications and regulations on aircraft impact on airports and on ventilation systems in buildings or transportation systems (cars, trains, or airplanes), which in turn impose lower and lower maximum noise levels to such systems. For instance, turbofan engines have increased their bypass ratio in order to improve the aircraft performance while diminishing the nominal speed of rotation. The jet noise is then reduced, and the fan noise becomes a dominant source of noise, especially at approach. A quick calculation of the overall noise generated by a given fan geometry, either low speed or high speed, would be a valuable asset to any fan design designer and manufacturer prior to any installation in a building or an airplane, for instance. However, an accurate prediction of the sound by any full turbomachinery still remains a challenging goal and a daunting task to be achieved by a direct computation. In the present study, the noise predictions will then rely on a strip theory combined with an acoustic analogy based on the wall pressure fluctuations. For the low-speed fans, the model is an extension of the development by Schlinker & Amiet for helicopters. Turbofan engines induce two additional difficulties to noise modeling: the cascade effects and the duct configuration which are presently modeled.

2:40

2pNSa5. Sound generation mechanism in low speed axial fans. Stefano Bianchi, Alessandro Corsini (Dept. of Mech. and Aeronautics, "Sapienza" Univ. of Roma, Via Eudossiana 18, 00184 Roma, Italy, bianchi@dma.ing.uniroma1.it), and Anthony G. Sheard (Flakt Woods Ltd., Colchester, Essex CO4 5ZD United Kingdom)

Because of the global civil regulation concerning acoustic emission, the goal of the manufacturers is to substantially decrease the noise radiated by the low-speed fans, without degrading their aerodynamic performance. One of the main goal to accomplish this target is the evaluation of the unsteady aerodynamic sources in the fan rotor responsible for the noise emission. An experimental analysis, based on a cause-effect method, was carried out and the results are presented in this paper. The process focuses on the experimental comparison between the novel prototypes of a family of low-speed fans and enabled the evaluation of the noise sources in the rotor near-field domain. The near-field aerodynamic sources accounted in this work correspond to the zones with high pressure fluctuations amplitude, located in the wake region of the exhaust flow. The sources were measured along the rotor radius and then correlated with the noise emitted by the fan at the far-field domain, using a Fourier based methodology. The analysis was carried out looking at the cross-spectra polar form using the phase shift as revealing tool for aerodynamic noise sources. The results were compared to evaluate the merit of the aero-acoustic performance of each prototype.

3:00—3:15 Break

3:15

2pNSa6. Quiet, efficient fans for space flight: An overview of NASA's technology development plan. Danielle Koch (NASA Glenn Res. Ctr., 21000 Brookpark Rd., MS 54-3, Cleveland, OH 44135, l.danielle.koch @nasa.gov)

NASA has developed a Technology Development Plan for improving the aerodynamic and acoustic performance of spaceflight fans. The intent is to make broader use of the technology developed at NASA Glenn to improve the efficiency and reduce the noise of aircraft engine fans. The goal is to develop a set of well-characterized government-owned fans nominally suited for spacecraft ventilation and cooling systems. NASA's Exploration Life Support community will identify design point conditions for the fans in this study. Computational Fluid Dynamics codes will be used in the design and analysis process. The fans will be built and used in a series of tests. Data from aerodynamic and acoustic performance tests will be used to validate performance predictions. These performance maps will also be entered into a database to help spaceflight fan system developers make informed design choices. Velocity measurements downstream of fan rotor blades and stator vanes will also be collected and used for code validation. Details of the fan design, analysis, and testing will be publicly reported. With access to fan geometry and test data, the small fan industry can independently evaluate design and analysis methods and work toward improvement.

3:30

2pNSa7. Aerodynamic design and computational analysis of a space flight vehicle cabin ventilation fan. Daniel L. Tweedt (AP Solutions, Inc., 22526 Tammy Circle, Council Bluffs, IA 51503, daniel.tweedt@ap-solutions.com)

Quieter working environments for astronauts are needed if future long-duration space exploration missions are to be safe and productive. Ventilation and payload cooling fans are known to be dominant sources of noise, with the International Space Station being an important case in point. To address this issue in a cost effective way, early attention to fan design, selection, and installation has been recommended. Toward that end, NASA has begun to investigate the potential for small-fan noise reduction through improvements in fan aerodynamic design. Using tools and methodologies similar to those employed by the aircraft engine industry, the aerodynamic design of a new cabin ventilation fan has been developed, and its aerodynamic performance predicted using computational fluid dynamics (CFD) codes. The design, intended to serve as a baseline for future work, is discussed along with selected CFD results.

3:45

2pNSa8. Surface integration method used on semi-solid surface for fan noise prediction. Yoon-Shik Shin and J. Stuart Bolton (Ray W. Herrick Labs., Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031, shin31@purdue.edu)

As a continuation of previous work, the Ffowc-Williams and Hawkings model was applied to an axial fan that was operated under unfavorable inflow condition. A thin aluminum foam screen that is used as a treatment for disturbed inflow condition was included as a noise source in computational aeroacoustic simulation. The incompressible URANS model was used for the CFD simulation and the inflow foam was modeled as a homogeneous porous zone with permeability and inertial resistance factors found experimentally. Since the homogeneous porous zone defined in this way does not have a solid structure that could be used as a dipole noise source, the control surface integration method was applied to the surface of the homogeneous porous zone: i.e., to a surface right between the normal fluid zone and the porous zone. The unsteady pressure data on the control surface under the incompressible assumption was treated as representing pressure fluctuations on a semi-solid surface rather than as a representation of acoustic sources within the volume surrounded by the control surface.

4:00

2pNSa9. Cooling fan noise reduction by optimizing spacing of blades. John Wang (Volvo Construction Equip., 312 Volvo Way, Shippensburg, PA 17257, john.wang@volvo.com)

Cooling fan noise theories are briefly reviewed. A simplified model is derived from a rotating point source theory. Optimization was done using the simplified model. A fan with unevenly spaced blades was prototyped, tested, and compared with a fan with evenly spaced blades. Noise reduction data are presented.

4:15

2pNSa10. Suppression of ducted dipole noise by an expanded segment around the source. Lixi Huang (Dept. of Mech. Eng., Univ. of Hong Kong, Hong Kong, lixi@hku.hk)

Low-speed fan noise mainly features dipole sound radiation. For such a fan operating in a duct, the low- to medium-frequency radiation derives mainly from the unsteady, axial forces acting on the rotating blades. Recently, we have found, both theoretically and experimentally, that self-cancellation of sound can be achieved by placing the dipole in an expanded segment of the duct which provides a reverberating environment with a reduced radiation impedance. The insertion loss spectrum shows a very broadband appearance similar to but is better than the transmission loss spectrum of an expansion chamber silencers used for reflecting an incident plane wave. Parametric analysis reveals that the expansion ratio controls the peak insertion loss, while the chamber length controls the single-lobe bandwidth. Simple plane wave analysis implies that short chamber would give wider stopband with a chamber of vanishing length yielding the best result. It is shown here that such analysis fails when the chamber is too short. For a given finite frequency band, there exists an optimal chamber length. Spectral element analysis reveals details of such optimal length and the result is explained in terms of the known fundamentals of dipole radiation and duct acoustics.

4:30

2pNSa11. Active control of axial fan noise using a multiple-input multiple-output feedback controller. Cole V. Duke, Scott D. Sommerfeldt, and Kent L. Gee (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602, coleduke@gmail.com)

In the past, significant progress has been made in actively controlling the tonal noise of axial cooling fans using a digital, feed-forward controller. Progress has also been made in controlling broadband fan noise using an analog feedback controller due to the nondeterministic nature of the noise. Current work focuses on the control of broadband noise with a multiple-input multiple-output (MIMO) feedback controller using four control sources and four error sensors. To ensure stability of the closed-loop control system, the phase and magnitude characteristics of the controller must be chosen to produce attenuation in the target frequency band without augmenting the noise outside the band. A practical analog feedback controller will be presented, and performance results will be compared between a single-input single-output controller and a MIMO controller.

Session 2pNSb

Noise, Animal Bioacoustics and INCE: Effects of Noise on Humans and Non-Human Animals II

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Einsteinufer 25, Secr TA 7, Berlin, 10587, Germany

Invited Papers

1:05

2pNSb1. Determining the effects of long-range missile noise on wildlife. Timothy Lavallee (LPES, Inc., 14053 Lawnes Creek Rd., Smithfield, VA 23430, tlavallee@lpesinc.com)

This is presentation on determining the sound levels from missile test and the potential effects on wildlife. The discussion will include an estimate noise from the launch, rocket motor splashdown, sonic boom, and detonation of long-range missiles. An overview of calculations and modeling to determine exposure to wildlife will be incorporated. The discussion will include overview of the ongoing development of noise threshold criteria for marine mammals and terrestrial wildlife, particularly, dolphins, pinnipeds, and avian species. A literature review and detailed review of the effects threshold will be included. A discussion of the benefits and the shortcoming associated with both the noise modeling and the limitations with respect to the existing effects thresholds will be included.

1:25

2pNSb2. Addressing concerns about ultrasound noise during expansion of a laboratory animal facility. Gary M. Glickman (Wilson, Ihrig & Assoc., Inc., 65 Broadway, Ste. 401, New York, NY 10006, gglickman@wiai.com)

Since 2002 the mouse has become the preferred model for biomedical research largely due to the ease with which it can be genetically engineered. The growth in genetic research has led many institutions to expand their animal holding and research facilities. Managing expectations of investigators during the process of construction and minimizing potential for noise impacts to research are critical for a successful expansion project. Noise in the ultrasound frequency range is often a concern among researchers due to limited data on its causes and response of mice to it and due to the fact that humans cannot perceive ultrasound and are therefore unable to detect when it occurs. Consequently, prior to and during construction, issues involving ultrasound are best given careful consideration so as to minimize it and at the same time address researcher's concerns. Work will be presented from a case study of an expansion project at a medical research facility involving transgenic mice. Issues surrounding the potential for construction noise to impact mice holding rooms will be examined. Particular attention will be paid to measurement and evaluation of noise at ultrasound frequencies and approaches for addressing the concerns of investigators.

Contributed Papers

1:45

2pNSb3. Animal magnetic resonance imaging and acoustic noise. Amanda M. Lauer (Dept. of Otolaryngol.-HNS, Johns Hopkins Univ., 521 Traylor Bldg., 720 Rutland Ave., Baltimore, MD 21205), AbdEl-Monem M. El-Sharkawy, Dara L. Kraitchman, and William A. Edelstein (Johns Hopkins School of Medicine, Baltimore, MD 21205)

MRI scanner noise can cause discomfort and potential damage to human hearing. MRI is used extensively in animal research and, increasingly, as a diagnostic tool in veterinary medicine, generally with no hearing protection. Effects of noise could be detrimental for research animals that will be used after scanning or for companion animals that will be returned to owners. Though safe exposure standards have not been determined for most animals, it is important to consider that animals exposed to scanner noise may suffer hearing damage. To study potential animal MRI adverse hearing effects, we measured sound levels produced by several animal scanning protocols in a commercial 3T scanner. Using hearing threshold data for research or companion animals, we estimate weighted sound pressure levels and compare these to levels at which damage occurs in humans. SPLs were above 90 dB on many scans and often exceeded 100 dB. Animals may be exposed to these levels intermittently for 1 h or more. Exposure to these sound levels could result in temporary or permanent hearing loss. Thus, it is important to use hearing protection, quieter pulse sequences, or quieter scanners for animal research or companion animal veterinary MRI.

2:00

2pNSb4. Comparison of impulse noise damage risk criteria using the chinchilla impulse noise exposures. William J. Murphy, Amir Khan (Hearing Loss Prevention Team, CDC/NIOSH, 4676 Columbia Parkway, MS C-27, Cincinnati, OH 45226, wjm4@cdc.gov), and Peter B. Shaw (CDC/NIOSH, Cincinnati, OH 45226)

The 1968 CHABA recommendations to limit impulsive noise exposure to levels below 140 dB sound pressure level form the basis of current United States occupational and military standards. The U.S. military standard, MIL-STD 1474D, estimates the number of allowable shots to which a person may be exposed using peak level, B-duration, for varying levels of hearing protection usage. The French Weapons Noise Committee has uses the 85 dBA *A*-weighted equivalent level, $L_{\mathrm{Aeq8\ hr}}$, as the limit for allowable exposures.

The U.S. Army sponsored a series of noise exposures with chinchillas to investigate the effects of level, number of impulses, and interstimulus interval. Several types of impulses were created ranging from acoustic shock tubes to narrow band impacts reproduced by a loudspeaker. The goodness-of-fit and the discrimination of five noise exposure metrics were evaluated in this study: MIL-STD 1474D, AHAAH model, $L_{\mathrm{Aeq8\ hr}}$, Pfander's C-duration metric, and Smoorenburg's D-duration metric. Goodness-of-fit was evaluated with a logistic regression and discrimination was evaluated using the area under the receiver operator characteristic curve. The $L_{\mathrm{Aeq8\ hr}}$ was found to best predict the temporary threshold shifts and the AHAAH model was found to best predict the permanent threshold shifts. [Partial funding provided by US Army Aeromedical Research Laboratories MIPR8J07586218].

2:15

2pNSb5. Extraction of hearing loss threshold levels and equal auditory risk metric curves of chinchillas from existing exposure data. Wonjoon Song, Steve Goley, and Jay Kim (Dept. of Mech. Eng., Univ. of Cincinnati, Cincinnati, OH 45221-0072)

An existing set of chinchilla noise exposure data has been re-analyzed using an advanced signal analysis tool and a statistical regression analysis. The data are comprised of 23 digitally recorded noise files and auditory damage indicators of the chinchillas measured at six frequency points, 0.5, 1, 2, 4, 8, and 16 kHz. From the relationship between the equivalent sound pressure level (Leq) of the six 1/3 octave frequency components of the noise and PTS of chinchillas, the Leq level that starts to cause auditory damage to chinchillas is identified as a function of frequency. These threshold levels of chinchillas, obtained for the first time in this study, will be useful information to design chinchilla exposure studies. The levels that will cause the same amount of PTS at the six frequencies are also obtained, from which equal auditory risk metric (EARM) curves are constructed. It is shown that a chinchilla version of an advanced noise guideline which enables quantitative, frequency-by-frequency assessment of the risk of a noise can be developed using the EARM curves. Possibility as well as difficulties of developing a human version guideline similar to this chinchilla prototype are discussed.

2:30

2pNSb6. Survey of ambient noise in aquariums. Colin W. Jemmott (Penn State Grad. Program in Acoust., P.O. Box 30, State College, PA 16804, cwj112@psu.edu)

Owning and maintaining an aquarium is a common hobby, but the ambient noise in aquariums resulting from pumps, filters, bubblers, and other equipment is not well studied. Elevated ambient noise levels have been shown to adversely affect wild fish and marine invertebrates, and anecdotal evidence suggests that this may be a problem in aquariums as well. Aquariums designed to maintain coral reefs require high water flow and pristine water conditions, which in turn require pumps and protein skimmers that contribute to underwater noise. A survey of ambient noise in both fresh and saltwater aquariums ranging in size from 5 to 550 gal was conducted. The aquariums differed in construction material, number, size and type of pumps, and presence of other equipment. The ambient noise broadband levels, spectrum levels, and time-frequency representations are compared, and general conclusions are drawn about noise characteristics of equipment and aquariums.

2:45

2pNSb7. A new tranquility rating tool and the role of audio and visual interactions. Kirill Horoshenkov, Robert Pheasant, Greg Watts (School of Eng., Univ. of Bradford, Bradford BD7 1DP, United Kingdom, k.horoshenkov@bradford.ac.uk), David Whitaker (Univ. of Bradford, Bradford BD7 1DP, United Kingdom), and Jian Kang (Univ. of Sheffield, Sheffield S10 1FL, United Kingdom)

Restorative environments which enable individuals to recover the sense of well-being are becoming increasingly important. These environments are characterized by an enhanced level of tranquility. Therefore, obtaining the optimum balance between the landscape and soundscape characteristics and measuring the resultant tranquility of these spaces are essential for their design and maintenance. In order to understand the key factors which affect the tranquillity construct, a large volume of audio and visual data has been collected across a representative range of landscapes in the UK. These data have been analyzed objectively by studying the temporal and spectral characteristics of the recorded sounds and the proportion of the natural and contextual features present in the video clips. The tranquility rating of these landscapes has been obtained from subjective experiments on 44 subjects to whom uni- and bimodal stimuli have been presented in a separate experiment. The results of these experiments make possible to objectively measure the key components of the tranquillity construct and determine their relative importance in the design of a restorative space. On this basis new relations between the tranquility rating, sound pressure level, loudness characteristics, and the visual quality of the scene have been derived.

3:00

2pNSb8. The effect of anthropogenic noise on the calling behavior of amphibians and birds in urban areas of Puerto Rico. Maria Isabel Herrera-Montes and Mitchell Aide (Dept. of Biology, Univ. of Puerto Rico, P.O. Box 70377, San Juan, PR 00936-8377, isahemontes@yahoo.com)

The bird and amphibian communities of Puerto Rico have high levels of diversity and endemism. Although Puerto Rico has a very high density of roads and high levels of noise pollution, presently 40% of the island is covered in forest. I will address the following questions: Do high levels of anthropogenic noise change species composition of amphibians and birds? Are amphibians and birds modifying the time of calling and/or characteristics of their vocalization in response to anthropogenic noise? I collected sound recordings in two habitats. For recordings I used the automated recording devices developed by our research group. Each recording device was programmed to record for 7 consecutive days; 1 min every 20 min. Previous results showed that noise influences the bird community composition in secondary lowland forest sites, but not in karst forest. Another possibility is that bird and amphibian use other strategies to minimize the effect of noise. This can involve evolutionary changes in signal characteristics, as a long term adaptation or the species may be adjusting signal traits in response to variations in noise levels as a short term adaptation. I will answer these questions when I do analyzing all data.

3:15—3:45 Panel Discussion

Session 2pNSc

Noise, Animal Bioacoustics, and INCE: Acoustics and Public Policy

Nancy S. Timmerman, Chair
25 Upton St., Boston, MA 02118

Invited Papers

1:20

2pNSc1. Hearing protection devices: The regulatory journey to change. Catrice Jefferson (U.S. Environ. Protection Agency, 1200 Pennsylvania Ave., NW, Washington, DC 20460)

The regulation for the labeling of hearing protection devices (40 CFR Part 211, Subpart B) was promulgated in 1979 and has been the directive for 30 years. During this time, there have been significant changes in the industry with respect to newly developed products, which has resulted in the demand for various types of hearing protectors and an increased awareness of hearing loss in both occupational and non-occupational environments. The foundation of the newly proposed regulation of August 5, 2009 is based on these observations and those that were presented by industry, hearing conservation professionals, and the general public during the EPA Workshop on Hearing Protection Devices that was held in March 2003. While these observations will serve as the backdrop of the presentation, the scenery will be the key elements of the proposed regulation, such as the testing methodology, noise reduction rating range, label format and content, recurrent testing, and reporting requirements of testing data. The audience will journey through these scenic elements but will be delivered to its final destination, which are regulatory change and the prospects of withstanding another 30 years.

1:40

2pNSc2. Proposal for assessing the effectiveness of noise policies. Lawrence Finegold and Michiko So Finegold (Finegold & So, Consultants, 1167 Bournemouth Ct., Ctr.ville, OH 45459, lsfinegold@earthlink.net)

During the last half of the 20th century, most of the developed countries of the world have been addressing noise as an environmental and occupational problem, and have been working to develop noise control and exposure policies and related noise control technologies. Since the 1970s, the United States has been an international leader in these efforts, although more recently significant advances have been made elsewhere, particularly by the European Commission. However, little is known about how effective various noise policies and regulations have been in controlling the noise exposure of the populations which they are intended to protect. This paper proposes a national program to address this issue. The first phase involves the collection and cataloging of noise policy regulations and related guidelines and, for the most part, has already been completed by the International Institute of Noise Control Engineering, including a comprehensive index of U.S. noise policies. The second phase would involve developing a baseline of noise exposure estimates for various community environments. The third phase would determine the long-term effectiveness of noise policies in controlling noise exposure by examining the changes in various noise exposures and noise effects over time.

2:00

2pNSc3. Integrating disturbance ecology and public policy: A better framework for managing the effects of anthropogenic noise. Ann E. Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109) and Amy R. Scholik-Schlomer (Natl. Marine Fisheries Service, Office of Protected Resources, 1315 East-West Hwy., Silver Spring, MD 20910)

The Marine Mammal Protection Act and the Endangered Species Act are statutes designed to conserve and recover marine mammal species. In addition to language that restricts activities that could "take" individuals via death, they restrict activities that could harass marine mammals. By framing the problem in terms of harassment, these statutes acknowledge the potential for significant behavioral effects. However, predicting biologically significant effects of disturbance is often difficult. A better theoretical and technical framework is needed that accounts for the adaptive scope (both physiological and behavioral) of large, intelligent, wide-ranging, and long-lived marine mammals. While ASA members have debated policy statements on the inappropriate use of acoustic concepts in controversies over marine mammal exposure to noise, they have not effectively addressed the need for better integration of biological theory. Of particular importance are theories on cognitive function, adaptive defensive strategies, and allostasis that help explain why marine mammal responses can be intelligent and flexible in some cases and injurious in others. By integrating such biological concepts into its policy statements, the Society could foster more effective use of research resources and help stakeholders support regulatory instruments for noise that would effectively balance human needs with those of marine mammals.

2:20

2pNSc4. Disastrous decisions: Costs to natural quiet of Grand Canyon National Park. Dickson J. Hingson (275 S. River Run, Flagstaff, AZ 86001)

Noisy intrusions from air tour rides and other aircraft increasingly, adversely affect highly significant areas of the Grand Canyon National Park. These continue in the heart of the park, even though Theodore Roosevelt told us to "do nothing to mar its grandeur." Preserving the Canyon's numinous aura requires maintaining its emblematic silence. The requisite policy would be to severely restrict aircraft noise. Belated efforts for an aggressive noise policy began here soon after 1965 airport completion near the Park. Hard won Congressional mandates for restoration of natural quiet were nonetheless frustrated by an accelerating aircraft boom fed by still more bad decisions, FAA stonewalling/bullying, and industry selfishness. The NPS/FAA environmental outcome—being finalized now, under a "change" administration—is coming into focus via a detailed EIS under public comment. The still unmet national direction to curb this

noisy, commercial park "take-over" illustrates serial failures in group decision-making. A "road map" of contributing reasons for a veritable acoustic/soundscape "collapse" here gains strength from insights in Jared Diamond's recent book "Collapse":, i.e., failures to anticipate, to perceive, and to behave or obey. To right these wrongs, the concerned public must assertively guide the current Administration's implementation of substantial, enduring restoration of the Park's natural quiet.

2:40

2pNSc5. Sex, money, religion, and politics: Understanding noise policy in terms of what is left unsaid. Leslie D. Blomberg (Noise Pollution Clearinghouse, Box 1137, Montpelier, VT 05601-1137)

Sex, money, religion, and politics. We are not supposed to talk about these things. And when it comes to discussions of creating a comprehensive noise policy, we generally do not, and perhaps we should not. But we should at least understand their contribution to the mxe9;lange of laws and regulations that is our current noise policy. This paper examines the elephant in the room that people are not talking about.

3:00—3:20 Break

3:20

2pNSc6. Would Western noise policies be effective in developing countries? Lawrence Finegold (Finegold & So, Consultants, 1167 Bournemouth Ct, Ctr.ville, OH 45459, Lsfinegold@earthlink.net) and Dieter Schwela (Swedish Environ. Inst.)

This paper examines the question of whether current and evolving Western noise policies would be effective and appropriate for use in developing countries. Differences in noise sources, available finances and noise control technologies, cultural norms, climate, views concerning the role of the government, etc., make it likely that different approaches might be needed in developing countries for their noise policies to be effective. Thus, it recommends that an international consortium of scientists and engineers, government representatives, and key stakeholders be organized to address this important topic. The noise policy needs of developing countries require an awareness of their typical noise problems, along with appropriate, affordable, and technologically feasible solutions, and how these might be incorporated into emerging Global Noise Policy concepts. This paper presents a concept for a Strategic Approach to Environmental Noise Management in Developing Countries, developed by the Swedish Environmental Institute (SEI), to form a foundation for the proposed international consortium. The next step is to now formalize an international consortium to provide additional technical input, to encourage participation by international agencies, professional societies and stakeholders, to organize workshops and special sessions at acoustics congresses, and to promote the evolving SEI concept with the governments of developing countries.

3:40

2pNSc7. There ought to be a law! Nancy S. Timmerman (25 Upton St., Boston, MA 02118, nancy.timmerman@alum.mit.edu)

This paper will examine some of the logical consequences of laws and policymaking as they relate to acoustics. In acoustics and noise control, "public policy" is used to try to legislate good behavior. Those who behave well do not need the laws. Those who do not will not obey the laws. Examples will be drawn from aviation and planning, examining the use of measures such as DNL. In particular, this paper will address how laws create unintended results, frequently not addressing the issue at all.

Contributed Papers

4:00

2pNSc8. New York City: The (E) designation and restrictive declaration. Benjamin Sachwald (AKRF, 440 Park Ave., South 7th Fl., New York, NY 10016, bsachwald@akrf.com)

Urban areas are typically associated with high-noise levels. With a population of more than 8×10^6 and growing, New York City is one of the largest cities in the world. In New York City, because of the population density and the wide variety of land uses (for example, residential, commercial, transportation, manufacturing, and industrial) that exist in close proximity, it is common for a residential building to be located in an area with ambient noise levels that would be considered unsuitable for residential use. To protect inhabitants of a planned building to be located in an area with high ambient noise levels, an (E) designation or a Restrictive Declaration may be used to ensure that the building's interior environment meet a certain acoustical design criterion. In this paper, the (E) designation and Restrictive Declaration are defined, their regulatory process is outlined, and specific case studies are discussed.

4:15

2pNSc9. Recommendations for the improvement of the management of noise from helicopter operations. Paul Kendrick, David C. Waddington, Geoff Kerry (School of Computing, Sci. and Eng., Univ. of Salford, Salford M5 4WT, United Kingdom, p.kendrick@salford.ac.uk), Matthew Muirhead, and Ray Browne (QinetiQ Ltd., Farnborough GU14 0LX, United Kingdowm)

This paper results from recent research commissioned by the Department for Environment, Food and Rural Affairs (Defra, UK) with the objective of identifying possible opportunities for the improvement of the management of noise from helicopter operations in the UK. Effected populations are not just those living close to heliports but include those exposed to noise from helicopters used by the military, emergency services, and commercial companies. A detailed report has been produced addressing the improvement of the management of helicopter noise. Also produced was a short nontechnical guide including the means of redress for perceived disturbance. The report also makes clear recommendations for future work and identifies particular areas where specific research is required. This paper discusses these recommendations and suggests a way forward for research into the improvement of the management of helicopter noise. [Work funded by the Department for Environment, Food and Rural Affairs (Defra), UK].

Session 2pPA

Physical Acoustics and Engineering Acoustics: Ultrasonics, Nonlinear Acoustics, Acousto-Optics, and Engineering Acoustics in Honor of Mack Breazeale II

Laszlo Adler, Cochair
Adler Consultants Inc., 1275 Kinnear Rd., Columbus, OH 43212

Igor Ostrovskii, Cochair
Univ. of Mississippi, NCPA, Dept. of Physics, University, MS 38677

Chair's Introduction—12:55

Invited Papers

1:00

2pPA1. Highlighting the career of Mack Breazeale. James G. Miller (Dept. of Phys., Washington Univ., 1 Brookings Dr., St. Louis, MO 63130, james.g.miller@wustl.edu)

Mack Breazeale's creative work was respected and recognized by professional organizations including the Acoustical Society of America, the Institute of Acoustics (Great Britain), and the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society (UFFC), each of which honored him with the rank of Fellow. Breazeale served as the UFFC Distinguished Lecturer in 1987–1988, during which time he traveled extensively to deliver highly regarded presentations. In recognition of Breazeale's many contributions to the IEEE UFFC, including his many-year participation in the Technical Program Committee that sets the theme for the Annual IEEE Ultrasonics Symposium, Breazeale was honored with the Achievement Award in 2008, the highest Society-wide award presented to a member in special recognition of outstanding contributions. I had the pleasure of highlighting Breazeale's many contributions that led to this prestigious honor during the awards ceremony at the 2008 Ultrasonics Symposium in Beijing. When I wrote to Breazeale to express my congratulations on this award and to discuss how I might summarize his many achievements in a career spanning decades, Breazeale responded by saying, "The most important aspect of my career is the fact that I have served as Major Professor to 15 MS candidates and 16 doctoral candidates."

1:10

2pPA2. Training graduate students: The teaching philosophy of Mack Breazeale. Michael S. McPherson (Dept. of Phys. and Astronomy, 1906 College Heights Blvd. #11077, Bowling Green, KY 42101, mike.mcpherson@wku.edu)

Mack Breazeale's last Ph.D. student will comment on Dr. Breazeale's contributions to solid state acoustics, nonlinear acoustics, and acousto-optics. In particular, this talk will focus on Dr. Breazeale's approach to, and methodology for, training and teaching graduate students how to conduct research and mature into capable scientists. This talk will speak to Dr. Breazeale's unique gifts as a teacher and mentor and his approach to conducting scientific work in general.

1:20

2pPA3. Research on acoustical memory discovered in the laboratory of Mack Breazeale. Igor Ostrovskii (Dept. of Phys., Univ. of Mississippi and NCPA, University, MS 38677, iostrov@phy.olemiss.edu)

Research at Mack's Laboratory at NCPA has discovered that a $LiNbO_3$ single crystal can memorize an ultrasonic pulse for tens of microseconds, the frequency range was 17–28 MHz. Supposedly, the Acoustical Memory (AM) effect is connected to the oxide internal microstructures. This conclusion is based on frequency, thermal, and ultrasound attenuation data. Mack Breazeale's leadership, scholarly character, and gentle nature were clearly revealed through these experimental investigations, various meetings, and scientific discussions in his office and laboratory. However, a basic connection between the ceramic microstructures and AM formation is not revealed yet. Further development of AM investigations is done in the inversely poled ferroelectric wafers of $LiNbO_3$ and $LiTaO_3$. The speeds of different acoustic modes are altered in the presence of periodically poled ferroelectric domains. Calculations for a bulk periodically poled $LiNbO_3$ yield a similar result despite the uniform acoustic properties along the specimen. Distortions in the dispersion curves of ultrasound in non-homogeneous ferroelectrics with specific internal microstructures may be a physical cause of acoustical memory. Some applications are discussed. [Work supported by UM.]

1:40

2pPA4. Theoretical and experimental investigation of the backward beam displacement. Nico F. Declercq (UMI Georgia Tech, CNRS 2958, 2 Rue Marconi, 57070 Metz France and Woodruff School of Mech. Eng., Georgia Tech, Atlanta, GA 30332-0405)

The discovery of a backward beam displacement of ultrasound interacting with a periodically corrugated surface, dates back from 1976, when Breazeale and Torbett reported it [Appl. Phys. Lett. **29**, 456 (1976)]. Since 2002 new investigations have been undertaken partially motivated by Breazeales enthusiasm. An overview is presented of how the phenomenon was first discovered in 1976, how a theoretical explanation was found since 2002 [Declercq *et al.*, Appl. Phys. Lett. **82**, 2533 (2003); Declercq *et al.*, J. Appl. Phys. **96**,

6869–6877 (2004)] and what further verifications and discoveries have been made since then [Teklu *et al.*, J. Appl. Phys. **97**, 084904 1 (2005)]. The report also highlights new research showing the ubiquitous presence of the phenomenon and its importance for nondestructive applications [Herbison *et al.*, J. Acoust. Soc. Am. In press].

2:00

2pPA5. Breazeale legacy in Gaussian acoustics. Dehua Huang (NUWC, Newport, RI 02841)

Professor Breazeale is a great educator, an outstanding researcher, and scientist in many acoustic fields. This talk focuses professor efforts and leadership in pioneer Gaussian acoustics theory and practice. In 1992, instead of going by KZK or Helmholtz equations for theoretical Gaussian acoustic wave solution, a new path was discovered based on the far field approximation from the Rayleigh integral. By which, the mathematical solution matches with the other Gaussian wave equation results. Further, the design of the Gaussian acoustic transducer technologies is evolved. It was only possible for 2 MHz and above frequency band operating Gaussian transducers. This frequency band barrier was crossed over, and a set of kilohertz Gaussian transducers was fabricated in the laboratory by a new design. The definition for a qualified Gaussian transducer was also finalized. The propagation of a Gaussian wave in an ocean environment was also simulated.

2:20

2pPA6. Acoustic parametric oscillators: In the footsteps of Mack Breazeale. Lev A. Ostrovsky (Zel Technologies and NOAA Earth System Res. Lab., 325 Broadway, Boulder, CO 80305)

Parametric oscillators have been studied in physics since the 18th century. In acoustics, the effect of transformation of pump energy to a half-frequency subharmonic (or, more generally, to any lower-frequency signals) began to be studied in 1960s by Korpel and Adler and in the laboratory of Breazeale. They observed basic parametric effects such as half-frequency generation and generation of lower-frequency doublets. In early 1970s, our group in Russia began studies of acoustic parametric phenomena in dispersive systems (waveguide resonators), in which most of extra harmonics and subharmonics are suppressed so that generation at the desired frequencies occurs much more efficiently. Indeed, in these experiments the subharmonic mode was generated so intensively that it can even damp the pumping mode in resonance. Also the later experiments with "non-classical," strongly nonlinear solid materials, such as ferroelectrics and metals, should be mentioned. The more recent works include parametric phenomena in bubbles; moreover, the group of Lauterborn in Germany has demonstrated that the period-doubling scenario of transition to chaos can be realized in cavitation noise. This presentation outlines a brief history of acoustic parametric oscillators and relevant effects in the context of pioneering works of Professor Mack Breazeale.

2:40

2pPA7. Close encounters of the distinguished kind. Koen Van Den Abeele (Wave Propagation and Signal Processing, K.U. Leuven Campus Kortrijk, E. Sabbelaan 53, B-8500 Kortrijk, Belgium, koen.vandenabeele@kuleuven-kortrijk.be)

Mack Breazeale entered my life at the very start of my research career in Kortrijk, Belgium. At the 20th anniversary of the university campus, he was co-organizer together with Oswald Leroy of the Physical Acoustics Meeting. I had the honor to edit and compile the book containing all contributed papers. Since that moment we have met each other regularly in Kortrijk or at an international meeting. Later he was a member of the jury at the defense of my doctoral thesis, and the same evening he managed to persuade my wife and me to come to NCPA for a year as a post-doc. We will always remember Ole-Miss and the parties at the gentle Breazeale family. Upon returning to Belgium I immediately got an email from Paul Johnson, triggered by Mack Braezeale, asking me to apply for a post-doc at Lanl. Thanks to Mack (and Paul), we had an unforgettable experience there. I consider it the birth of nonlinear acoustics and ultrasonics in Belgium. Oh, and by the way, we have performed some great science together in acousto-optics and nonlinear acoustics.

3:00—3:20 Break

Contributed Papers

3:20

2pPA8. Frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity. Alem A. Teklu (Dept. of Phys. & Astronomy, College of Charleston, 66 George St., Charleston, SC 29424, teklua@cofc.edu), Michael A. McPherson (Western Kentucky Univ., Bowling Green, KY 42101), Mack A. Breazeale (Natl. Ctr. for Physical Acoust., University, MS 38677), and Nico F. Declercq (Georgia Tech Lorraine, 57070 Metz-Technopole, France)

The excitation of fractional harmonics in a liquid-filled cavity by ultrasonic waves was described previously as a parametric phenomenon [L. Adler and M. A. Breazeale, J. Acoust. Soc. Am. **48**, 1077–1083 (1970)]. That is, by driving a transducer at one end of a fluid-filled cavity parallel to a rigid plane reflector at the other end, standing ultrasonic waves can be generated. Variations in the cavity length resulting from transducer motion lead to the generation of resonant frequencies lower than the drive frequency (known as fractional harmonics). The system was modeled by using a modified Mathieus equation whose solution resulted in the prediction of critical threshold drive amplitude for the excitation of parametric oscillation. The apparatus used by Adler and Breazeale was recently refined for accurate measurements of the threshold amplitude for parametric excitation at frequencies ranging from 2 to 7 MHz. The measurements showed that in this range the threshold amplitude increases with increasing drive frequency in apparent discrepancy with the results of Adler and Breazeale. Analysis of the theory indicates, however, that both past and current results lie in two different stability zones, and each is in agreement with the existing theory.

3:35

2pPA9. Optical Bragg imaging of acoustic fields after reflection. Michael S. McPherson (Dept. of Phys. and Astronomy, Western Kentucky Univ., 1906 College Heights Blvd. #11077, Bowling Green, KY 42101, mike.mcpherson@wku.edu), M. A. Breazeale (Univ. of Mississippi, University, MS 38677), Nico F. Declercq (Georgia Tech Lorraine, UMI Georgia Tech, 57070 Metz-Technopole, France), and Alem Teklu (College of Charleston, Charleston, SC 29424)

If visible light is incident at the Bragg angle on an acoustic wave of the appropriate frequency, the light will undergo Bragg diffraction from the acoustic wavefronts in a manner analogous to the Bragg diffraction of x-rays in a crystalline lattice. If the acoustic beam has previously passed through an object before having light diffracted from it, the structure of that object will be imaged in the optical diffraction orders. In traditional Bragg imaging applications, the sound is first transmitted through the object. In this presentation a new reflection-based method of Bragg imaging (where light is reflected from rather than transmitted through the object) will be discussed. This technique is capable of imaging structures at or near the surface of the object and may have potential industrial and biomedical NDT applications.

3:50

2pPA10. Infrasound-convection instability. Konstantin Naugolnykh (Dept. of Phys., Univ. of Colorado, 325 Broadway, Boulder, CO)

Heated below temperature stratified atmospheric layer is unstable and convection flow can be developed in such an area. Convection happens because warm less dense air goes up while cooler air comes down. Infrasound can initiate convection flow and being amplified by the flow. The process of infrasound-convection interaction is considered in the present paper. The model of convection in a compressible fluid is considered and analyzed. The conditions of infrasound amplification are obtained.

4:05

2pPA11. Ultrasonic imaging and characterization of objects submerged in highly attenuating fluids. Bart Raeymaekers, Cristian Pantea, Curtis F. Osterhoudt, and Dipen N. Sinha (Los Alamos Natl. Lab., MPA-11, P.O. Box 1663, MS D429, Los Alamos, NM 87545)

An ultrasonic scanning tool to image objects with high-spatial resolution submerged in highly attenuating fluids is described. The attenuation coefficient of the fluids used range over two orders of magnitude. For deeper penetration into the attenuating fluids, a parametric array with a difference frequency between 200 and 500 kHz was used. The primary frequencies that generated this difference frequency varied between 2 and 8 MHz. This parametric array provides a collimated beam for imaging. Both simple tone burst and frequency chirp signals were used for the measurement. Using this apparatus, the quality and accuracy of the image as a function of frequency, fluid attenuation coefficient, and fluid type (clear and suspensions) were investigated. It was found that imaging was possible over a wide range of attenuating coefficients and fluid types.

4:20

2pPA12. Batchelor's fourth order velocity correlation. John M. Seiner (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, University, MS 38677)

The Lighthill theory for aerodynamic noise relates far field sound pressure to complex fourth order two point two-time space-time velocity correlations. The Mani–Gliebe–Balsa (MGB) jet noise prediction code follows Lighthill's theory in the development of its algorithm. Since fourth order turbulence correlations are not modeled in the MGBs flow solver, it is necessary to resort to some form of approximation. The MGB invokes Batchelor's fourth order decomposition, which is exact for homogeneous and isotropic turbulence. For anistropic flow, like that of a jet, it is not known what the error is in adopting Batchelor's isotropic decomposition. The question here is can stereo particle image velocimetry (SPIV) provide a means to establish the error incurred with adoption of this decomposition. In an attempt to define the penalty for adopting the Batchelor's decomposition, a simple 2×94; diameter laboratory jet was operated at ambient temperatures at Mach 0.85. SPIV measurements were acquired of the fourth order two-point space correlation and all of the other second order correlations. A comparison was made to numerical plume simulations with an advanced numerical flow solver running in the RANS mode. This paper provides an explanation for observed variations between measured and predicted jet noise.

4:35

2pPA13. Multipole expansion for nonlinear acoustic scattering at the difference-frequency. Glauber T. Silva (Inst. of Phys., Univ. Fed. Alagoas, Maceio, Alagoas 57072-970, Brazil, glauber@pq.cnpq.br)

In this work, the scattering at the difference-frequency (DF) of two finite-amplitude plane waves by a rigid sphere in a fluid is studied. The DF-scattering is present in some acoustic imaging methods such as vibro-acoustography. This has motivated us to perform this study. We obtain for the first time the multipole expansion for the DF-scattered pressure in the farfield. Theory is based on the successive approximation and the Green's function methods. From the multipole expansion, we derive some quantities such as scattering cross-section and the total DF-scattered power. Furthermore, we show that the DF-scattered pressure increases with the observation distance r like $r \ln r$, while it varies almost linearly with the difference-frequency. Despite the amplification, we demonstrate that the DF-scattered pressure does not grow indefinitely because of absorption in the fluid. The theory is applied to the scattering in the Rayleigh limit (scatterer much smaller than the incident wavelengths). Only monopole and dipole moments are relevant in this limit. In conclusion, we believe that this theory can help to understand better the nonlinear scattering in acoustics.

4:50

2pPA14. A new form of Gaussian beam function as the base function for generic wave field analysis. J. J. Wen (VeroSound, Inc., 128 Powder Mill Rd., Sudbury, MA 01776)

Wave propagations are best described by three major approaches: the spherical wave approach represented by Rayleigh integral, the planar wave approach represented by spatial Fourier analysis, and the Gaussian beam approach. Spherical and planar wave approaches, both in form of double integrals, are analytically inconvenient and computationally expensive. In contrast, the Gaussian beam approach, with huge advantages in analytical operability and computational efficiency, has been quietly but firmly winning popularity. Dr. Breazeale's article: A diffraction beam field expressed as the superposition of Gaussian beams, co-authored with Wen, is not only a most representative work for Gaussian beam approach but also a most cited academic work in sound field analysis and computation. As a memory of Dr. Breazeale's pioneer role in the raise of Gaussian beam method, this article presents a new enhancement to the existing Gaussian beam method. A modified form of Gaussian beam function introduced as the base function set for wave field analysis greatly reduces the error rooted from the parabolic approximation, the number of Gaussian beams needed, and the run-time computation burden. The process for calculating the beam coefficients and beam parameters is also greatly simplified.

Session 2pPP

Psychological and Physiological Acoustics and Engineering Acoustics: Application of Auditory Models to Hearing Aids

Brent W. Edwards, Chair
Starkey Hearing Research Center, 2150 Shattuck Ave., Ste. 408, Berkeley, CA 94704

Invited Papers

1:00

2pPP1. The application of auditory models to hearing aid research. Brent W. Edwards (Starkey Hearing Res. Ctr., 2150 Shattuck Ave., Ste. 408, Berkeley, CA 94704, brent_edwards@starkey.com)

Audibility has been a focus of hearing aid research for much of the hearing aid field's existence. With modern hearing aids, audibility no longer limits performance for hearing aid wearers with moderate hearing impairment. Limitations in suprathreshold auditory-perception now need to be better understood in order to determine existing hearing deficits while aided. The ways in which hearing aids affect suprathreshold auditory function also need to be better understood in order to ensure that hearing aids provide maximal benefit in complex real-world situations. Auditory models can be used to address these needs. This talk will review current applications and future opportunities for the application of auditory models to hearing aid research.

1:20

2pPP2. Differentiating inner from outer hair cell damage: Measures and applications to model-based hearing aid design and fitting. Enrique A. Lopez-Poveda, Peter T. Johannesen, Patricia Prez-González, Jorge M. Mndez, and Almudena Eustaquio-Martín (Instituto de Neurociencias de Castilla y Leon, Univ. Salamanca, C/Pintor Fernando Gallego 1, 37007 Salamanca, Spain, ealopezpoveda @usal.es)

Hearing aids (HAs) are designed and fitted mainly to restore normal audibility and loudness growth. Unfortunately, after decades of research, many users still find them frustrating. We hypothesize that this is partly because HAs operate on the assumption that cochlear hearing loss is always associated with a reduction in cochlear gain and/or compression when this, in fact, need not be the case. Indeed, recent studies have shown that cochlear hearing loss frequently relates to selective or combined dysfunction of inner hair cells (IHCs) and outer hair cells (OHCs). Neither absolute thresholds nor loudness growth allows for a differential assessment of the type of hair cell damage, yet such assessment could be highly informative for optimizing HA design and fitting. In this presentation, we summarize our progress in designing a fast, universal method for inferring the degree of IHC and OHC dysfunction using distortion-product otoacoustic emission input/output functions. We also describe our progress in designing a HA based on these ideas and on a bank of dual-resonance non-linear filters inspired by computational models of healthy and impaired human cochlear responses. [Work supported by the MEC, Grant BFU2006-07536, and The Oticon Foundation.]

1:40

2pPP3. The use of a loudness model to derive initial fittings of multi-channel compression hearing aids. Brian C. J. Moore (Dept. of Experimental Psych., Univ. of Cambridge, Downing St., Cambridge CB3 9LG, United Kingdom, bcjm@cam.ac.uk)

A model of loudness perception applicable to hearing-impaired people has been used to develop three methods for the initial fitting of hearing aids with multi-channel compression, based on the audiometric thresholds. Method one, CAMEQ, has the goals of amplifying speech so that, on average, the specific loudness is the same for all frequencies within the range 500–4000 Hz, over a wide range of overall sound levels, giving about the same overall loudness as "normal" for speech covering a wide range of sound levels. Method two, CAMREST, determines the gains needed give normal specific loudness patterns for speech over a wide range of sound levels. Method three, CAMEQ2-HF, is similar to CAMEQ but differs from it in the following ways: (1) gains are recommended for center frequencies up to 10 kHz (as compared to 6 kHz for CAMEQ), (2) CAMEQ2-HF is based on the assumption that the user may wish to hear sounds from many directions and uses a diffuse-field-to-eardrum transfer function (as opposed to a free-field-to-eardrum transfer function for CAMEQ); (3) CAMEQ2-HF is based on recent wideband measurements of the average spectrum of speech. All of the methods can be applied to any multi-channel compression hearing aid.

2:00

2pPP4. Modeling of the auditory periphery for hearing aid evaluation and design. Ian C. Bruce (Dept. of Elec. & Comp. Eng., McMaster Univ., 1280 Main St. W., Hamilton, ON L8S 4K1, Canada, ibruce@ieee.org)

Amplification schemes and gain prescriptions for hearing aids have thus far been developed and evaluated based on perceptual criteria such as speech intelligibility, sound comfort, and loudness equalization. Finding amplification strategies that simultaneously optimize all of these perceptual metrics has proven difficult, despite decades of research. Furthermore, novel amplification schemes based on rough conceptual models of the normal and impaired auditory physiology have often proven to be unsuccessful. In this talk, I will describe studies directly employing physiologically accurate computational models to evaluate more rigorously what hearing aids are likely doing to the neural representation of speech. The results of these investigations indicate that (i) a physiologically accurate

auditory-nerve (AN) model can predict optimal linear and nonlinear amplification gains, (ii) optimal gains are dependent on both the spike-timing and mean-rate representations of speech in the AN, (iii) the proportion of outer hair cell and inner hair cell dysfunction in an impaired ear can affect optimal gains, and (iv) slow wide dynamic range compression (WDRC) or automatic gain control better restores the neural representation of speech than does fast WDRC.

2:20

2pPP5. Application of auditory models to assessing the sound quality of hearing aid signal processing algorithms. Volker Hohmann (Medizinische Physik, Universität Oldenburg, Oldenburg, Germany)

Advances in systems technology allow for increasingly complex processing algorithms in hearing systems addressing increasingly complex acoustic conditions. These developments have the potential of improving the rehabilitation of hearing impairment, but establishing reliable measures of benefit is quite difficult for these complex algorithms and conditions. Being the "gold standard" for algorithm evaluation, subjective testing of hearing-impaired subjects has some limitations in this context. It is very time consuming and the long acclimatization time needed when listening with new devices in complex acoustic conditions cannot easily be accounted for. Therefore, objective methods for estimating the benefit of an algorithm in a certain acoustic condition are desirable. They allow for identifying promising candidate algorithms and the acoustic conditions in which the algorithms might be applicable and thus identify critical acoustic conditions to be tested subjectively. In addition and in combination with technical measures such as segmental signal-to-noise ratio and distortion measures, perceptual measures based on auditory models might be useful for developing meaningful objective measures. This talk presents recent applications of auditory models to objective algorithm evaluation. Because of the growing importance of binaural and multi-microphone processing in hearing instruments, binaural/multichannel models for speech intelligibility and quality will be emphasized.

2:40

2pPP6. Auditory models of suprathreshold distortion in persons with sensorineural hearing loss. Ken W. Grant, Van Summers, Joshua G. W. Bernstein, Sandeep A. Phatak, Matthew J. Makashay, Elena Grassi, and Golbarg Mehraei (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307-5001, ken.w.grant@gmail.com)

Hearing technology has advanced to where it is now reasonable to ask whether signal processing algorithms can be developed to compensate for an individual's hearing loss, thus allowing them to hear functionally in a manner similar to persons with normal hearing. Clinically, the pure-tone audiogram is the primary tool used to represent the patient's hearing status. However, it has been well established, experimentally and theoretically, that the audiogram cannot reflect fully all aspects of the hearing loss, most notably that part which pertains to suprathreshold distortion. Much has been written about the distortion component of sensorineural hearing loss, yet there is little agreement on estimating its importance to speech recognition, nor much consensus on which hearing factors (e.g., spectral and/or temporal resolution) most accurately represent the distortion. Recent attempts to use biologically inspired models of auditory processing to represent a patient's internal auditory experiences have shown promise as a way to understand the role that suprathreshold distortions might play in speech recognition. This talk reviews recent work to develop auditory models of {\it individual} hearing-impaired listeners to predict differences in the perception of speech and non-speech signals not readily explained by audiometric thresholds. [Work supported by the Oticon Foundation.]

3:00—3:10 Break

3:10

2pPP7. Real-world application of auditory models to hearing instrument design. Stefan Launer (Phonak AG, Laubisruetistrasse 28, 8712 Staefa, Switzerland)

Auditory models provide a powerful tool for designing and developing signal processing algorithms of modern digital hearing instruments and can even be integrated directly into these algorithms. Models for loudness perception have been applied directly in hearing instrument algorithms with two applications: to control the gain of a hearing instrument and to limit the maximum output power of a hearing instrument as a "loudness limiter." Another example is the application of a very simple model in a hearing instrument to help resolve localization problems due to the microphone position in BTE hearing instruments, resulting in reduced front/back confusions. Examples of the application of models to the development and analysis of hearing instrument performance in real-life environments include the use of auditory scene analysis models to better understand human sound source classification and to improve the performance of automatic control systems in hearing instruments. A last example is the application of models describing human localization performance under realistic acoustic environments to understand the impact that hearing instrument signal processing has on the ability of hearing impaired people to localize sounds with hearing instruments. This talk will explore and discuss the potential of these different approaches for hearing instrument technology.

Contributed Papers

3:30

2pPP8. Moving fluid loading to the basilar membrane in passive nonlinear cochlear model. Svetlana Kovinskaya (Mechmath LLC, Prior Lake, MN 55372, mechmath@mechmath.com)

The mechanical response of the cochlea to sound is a nonlinear phenomenon. Existing models (with either passive or active outer hair cells) do not include effect of the motion of fluid loading to basilar membrane (BM). Being extremely slow this motion will play significant role only if the traveling wave in BM is close to the point of ceasing of propagation. Two fluid channels corresponding to the scalae vestibuli and tympani, each filled with incompressible liquid, provide the fluid loading to BM. Equation describing BM vibration with contra-directed moving fluid loading (from opposite sides of BM) becomes nonlinear equation because of an additional force proportional to loading velocity squared and the local curvature of BM (the second derivative of BM displacement). The velocity of this mass-like fluid motion is proportional to the applied sound power. In the case of sinusoidal excitation, the equation for BM vibration becomes the Mathieus equation with the characteristic equation determining stable and (parametrically) unstable regions. The nonlinear equation gives a multi-frequency quasi-steady state solution. Suggested mechanism (additional forces in the equation for BM vibration) explains the compression of strong stimuli, explains the generation of distortion products, and is in qualitative agreement with other experiments.

3:45

2pPP9. Predicting the effects of compression incorporating signal-adaptive band allocation. James M. Kates (GN ReSound, Dept. Speech Lang Hear Sci., Univ. of Colorado, 409 UCB, Boulder, CO 80309)

This paper considers the effects of dynamic-range compression on speech intelligibility and quality for hearing-impaired listeners. The paper focuses on a modification of compression in which the compression behavior for each frequency band is controlled by the peak amplitude within a frequency region surrounding the band. The stimuli are sentences in a background of multi-talker babble. Speech intelligibility and quality indices are used to model the listener responses for mild, moderate sloping, and moderate/severe hearing losses. The indices predict the effects of varying the compression parameters and the width of the frequency region, thus giving an indication of the expected benefit of the modified compression system and the optimal parameter settings as a function of hearing loss and listening conditions.

4:00

2pPP10. Prediction of binaural speech intelligibility when using non-linear hearing aids. Nicolas N. Ellaham (School of Inf. Tech. and Eng., Univ. of Ottawa, 800 King Edward Ave., Ottawa, ON K1N 6N5, Canada, nellaham@uottawa.ca), Christian Giguère (Univ. of Ottawa, Ottawa, ON K1H 8M5, Canada), and Wail Gueaieb (Univ. of Ottawa, Ottawa, ON, K1N 6N5, Canada)

A new objective measurement system is proposed to predict speech intelligibility in binaural listening conditions for use with hearing aids. Digital processing inside a hearing aid often involves non-linear operations such as clipping, compression, and noise reduction algorithms. Standard objective measures such as the articulation index, the speech intelligibility index (SII), and the speech transmission index have been developed for monaural listening. Binaural extensions of these measures have been proposed in the literature, essentially consisting of a binaural pre-processing stage followed by monaural intelligibility prediction using the better ear or the binaurally enhanced signal. In this work, a three-stage non-linear extension of the binaural SII approach is introduced consisting of (1) a stage to deal with non-linear processing based on a simple signal separation scheme to recover estimates of speech and noise signals at the output of hearing aids [Hagerman and Olofsson, Acust. Acta Acust. **90**, 356 (2004)], (2) a binaural processing stage using the equalization-cancellation model [Beutelmann and Brand, J. Acoust. Soc. Am. **120**, 331 (2006)], and (3) a stage for intelligibility prediction using the monaural SII [ANSI-S3.5, 1997 (R2007)]. Details of the new procedure will be discussed. [Research supported by NSERC (Canada).]

4:15

2pPP11. Predict the intelligibility of individual consonants in noise for hearing-impaired listeners. Feipeng Li, Woojae Han, and Jont Allen (Beckman Inst., Univ. of Illinois at Urbana, 405 N. Mathews Ave., Urbana, IL 61801)

The performance of hearing-impaired (HI) listeners to understand speech often deteriorates drastically in cocktail-party environments. To understand why a particular HI listener can hear certain sounds and not the others, it is critical that we take the prior knowledge about speech cues into consideration and investigate the effect of different types of cochlear hearing loss on speech perception. In the last 2 years we have tested over 50 hearing-impaired ears on consonant identification in noise. To evaluate the impact of shift of hearing threshold on the intelligibility of individual consonants in masking noise, a consonant intelligibility index (CII) is developed to predict the perceptual scores of individual consonants. Results of preliminary study indicate that CII makes an accurate prediction for flat mild-to-moderate hearing loss, but fails for the cases of cochlear dead region and extremely unbalanced (e.g., severe high-frequency) loss, suggesting that audibility alone does not fully account for the hard of speech perception in noise for many HI listeners.

4:30

2pPP12. Evaluating hearing aid processing with an auditory model of modulation sensitivity. Frederick J. Gallun (Natl. Ctr. for Rehabilitative Auditory Res., U.S. Veterans Hospital Rd., Portland, OR 97239, Frederick.Gallun@va.gov), Pamela E. Souza, and Eric Hoover (Northwestern Univ., Evanston, IL 60208)

Amplitude modulation is one of the fundamental aspects of sound to which humans are sensitive. A number of auditory models based on this sensitivity have been successful at explaining other aspects of human behavior, both in the domains of psychoacoustics and speech perception. A constrained-parameter model of perceptual similarity based on modulation sensitivity (the spectral correlation index) has recently been developed for use in the prediction of perceptual confusions at the phonemic level [Gallun and Souza (2008)]. The relatively small parameter set of the model allows for efficient adaptation of the basic architecture to novel problems. With a single change in the range of one parameter, this model can be used for the evaluation of hearing aid processing by predicting the confusions that will accompany various choices of amplification parameters. This approach demonstrates the value of evaluating signal processing strategies in hearing aids through changes in perceptual similarity as predicted by auditory models. [Work supported by NIH NIDCD R01 DC0060014 and VA RRD CDA-2 C4963W.]

4:45

2pPP13. Modeling quality judgments of music stimuli reproduced by a simulated hearing aid. James M. Kates (Dept. Speech Lang Hear Sci., GN ReSound, Univ. of Colorado, 409 UCB, Boulder, CO 80309) and Kathryn H. Arehart (Univ of Colorado, Boulder, CO 80309)

The focus in hearing aids is generally on speech, but the ability to reproduce music is also important for listener satisfaction. This study uses a sound quality metric to predict judgments of music quality made by normal-hearing and hearing-impaired listeners. A simulated hearing aid was used to process music modified by (1) additive noise and nonlinear processing, (2) linear filtering, and (3) combinations of noise, nonlinear processing, and linear filtering. The music stimuli included jazz, classical, and vocalize selections. The quality metric incorporates a model of the auditory periphery that reproduces aspects of impaired hearing. The effects of noise and distortion on the signal are estimated by measuring the changes to the envelope time-frequency modulation pattern, and the effects of linear filtering are estimated by measuring the changes to the signal long-term internal spectral representation. The accuracy of the quality metric for music will be presented, and the performance of the metric for music will be compared to its accuracy in reproducing the quality ratings of speech stimuli processed under similar conditions.

Session 2pSA

Structural Acoustics and Vibration: Computer Modeling of Structural Acoustics and Vibration for Complex Structures

James E. Phillips, Chair
Wilson Ihrig and Associates, Inc., 5776 Broadway, Oakland, CA 94618-1531

Chair's Introduction—1:25

Invited Papers

1:30

2pSA1. A multiphysics model of hearing aid vibroacoustics using finite elements. Thomas Burns and Dave Tourtelotte (Starkey Labs., Inc., Eden Prairie, MN 55344)

Electromagnetic balanced armature receivers [Hunt, *Electroacoustics*, Chap. 7] are used exclusively to generate acoustic output in hearing aids. These transducers are much more efficient than electrodynamic transducers and are capable of delivering upward of 140 dB of sound pressure to person. In an effort to maximize system gain in a hearing aid, vibroacoustical feedback paths originating from the receiver are modeled using finite elements. Given an electrical excitation, the electromagnetic-mechanical force on the armature is solved as a function of frequency. The force on this armature vibrates an internal diaphragm, which generates acoustic output while vibrating the entire hearing aid. Assuming that there are no acoustical leaks in the design, vibroacoustical coupling limits the usable gain of the aid. Using commercially available software, the fluid is modeled with full Navier–Stokes elements and is coupled to all structural boundaries. The armature is "kicked" with the aforementioned force, and spectral analysis is used to study mechanical transmissibility within and vibroacoustic reradiation around the hearing aid.

1:50

2pSA2. Finite element vibration analysis and measurement of a 255 ft vehicle ferry. Jesse H. Spence and Ron E. Dempsey (Noise Control Eng., 799 Middlesex Tpke., Billerica, MA 01821)

Finite element analysis (FEA) methods have become part of the standard repertoire when designing a ship. This is particularly true for static analyses, although FEA methods are becoming increasingly utilized for analysis of low-frequency vibration as well. This paper discusses the FEA modeling process that was used for analyzing low-frequency (less than 80 Hz) habitability vibration on a 255 ft vehicle ferry. Deck vibration levels were measured after the vessel was built. Measured versus predicted vibration data show good correlation, particularly when the various sources of potential error and uncertainty are considered. The subject ferry was designed to meet the ABS Comfort Class (COMF) vibration criteria. Calculations of deck vibration were performed using NEiNastran to verify compliance with the criteria. Excitations due to propeller hull pressure excitation are highlighted here as these forces created the largest deck responses. However, it is noted that the main propulsion engines were resiliently mounted subsequent to initial analyses using the FE model. This paper discusses pertinent details of the model (element types, properties, etc.), application of hull pressures, damping, analysis methodology, and interpretation of results. Areas of uncertainty and impacts to the results are discussed.

2:10

2pSA3. An evaluation of residual vectors in the commercial finite-element program NASTRAN. John B. Fahnline (Water Tunnel Bldg., University Park, PA, 16802)

MSC NASTRAN introduced residual vectors in 2005. They are used to improve solution accuracy by representing the static contributions of truncated modes in transient and frequency response calculations. This formulation is convenient because the residual vectors are added to the modal vectors to form an augmented basis set, where the residual vectors are orthogonal to each other and to the original mode set. Modal frequency response calculations then proceed as usual only with a few extra vectors in the basis set. This talk will focus on evaluating the improvements in accuracy that can be achieved using residual vectors for modal frequency response problems where two structures with truncated mode sets are joined together.

2:30

2pSA4. Derivation of a Ritz series modeling technique for acoustic cavity-structural systems based on a constrained Hamilton's principle. Jerry H. Ginsberg (School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, jerry.ginsberg@me.gatech.edu)

Hamilton's principle for dynamic systems is adapted to describe the fully coupled interaction of a confined acoustical domain and a bounding elastic system in cases of imposed velocity or pressure fluctuations on the surface or forced excitation of the structure. A key part of the modified principle is the treatment of the surface traction as a Lagrange multiplier function that enforces continuity conditions at the fluid-solid interface. The structural displacement, fluid velocity potential, and traction are represented by Ritz series, where the usage of the velocity potential as the state variable for the fluid assures that the flow is irrotational. Designation of the coefficients of the potential function series as generalized velocities leads to corresponding series representations of the particle velocity, displace-

ment, and pressure in the fluid, which in turn leads to descriptions of the mechanical energies and virtual work. Application of the calculus of variations to Hamilton's principle yields linear differential-algebraic equations whose form is identical to those governing mechanical systems that are subject to nonholonomic kinematic constraints. Criteria for selection of basis functions for the various Ritz series are illustrated with an example of a rectangular cavity bounded on one side by an elastic plate and conditions that change discontinuously on other sides.

2:50

2pSA5. Acoustic color of viscoelastic objects in two-fluid environments: High-fidelity, high-speed, three-dimensional finite-element modeling. David S. Burnett (Naval Surface Warfare Ctr. Panama City Div., 110 Vernon Ave., Panama City, FL 32407, david.s.burnett@navy.mil)

NSWC PCD has developed a computer simulation system for modeling the acoustic color (target strength versus frequency and aspect angle) of realistic three dimensional (3-D) objects that are near to or straddling the interfaces between different fluids. It employs high-fidelity (fully 3-D physics throughout object and environment), finite-element (FE) modeling of acoustic scattering from viscoelastic objects. It is implemented in a scalable-architecture, multiblade rack system that automatically runs hundreds of thousands of FE models, changing the mesh resolution and outer fluid boundaries of the models as they sweep over frequency. This frequency-dependent modeling yields a frequency-independent modeling error for the acoustic color response. This, in turn, yields higher accuracy for subsequent signal processing of the acoustic color data. The first part of this paper will present a brief overview of the theoretical, numerical, and software issues. The second part will present verification and validation of the system: (1) ultra-high-precision correlation with *T*-matrix simulations for four canonical target/environment configurations and (2) experimental validation for objects in free space and straddling the interface between two different fluids. [Work supported by ONR and SERDP.]

3:10—3:25 Break

3:25

2pSA6. Energy finite element analysis of naval vehicles. Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109, nickvl@umich.edu) and Kuangcheng Wu (Northrop Grumman Shipbuilding, Newport News, VA 23607)

Simulation methods can be used for meeting both the new acoustics and vibration initiatives for surface ship silencing. An energy finite element analysis (EFEA) formulation has been developed for computing efficiently the vibration of complex structures and the associated radiated noise at frequencies beyond the limits of conventional finite element methods. The EFEA provides a very attractive alternative to the statistical energy analysis (SEA) because it uses the same finite element model which is utilized for other simulations (i.e., structural analysis), while the SEA requires the development of a completely new model based on a lumped parameter approach. Further, the SEA requires estimates of modal densities for the various subsystems which introduce difficulties and approximations in the solution. Thus, the EFEA streamlines the structural-acoustic analysis within the ship design process due to modeling commonality with other disciplines. In this presentation, the main theoretical aspects of the EFEA will be presented first. Results presented for two 1/8th scale structures representing an advanced double hull design and a conventional hull design of a surface ship are analyzed. Results for the structural vibration induced on the outer bottom part of the structures are compared to available test data.

Contributed Papers

3:45

2pSA7. Engaging structural-acoustic simulations in multi-discipline design. Nickolas Vlahopoulos (Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48108, nickvl@umich.edu) and Jim He (Michigan Eng. Services, LLC, Ann Arbor, MI 48108)

In order to be effective and maximize the weight and cost savings, designing for noise and vibration attributes must be performed in parallel when designing a complex system for other disciplines (i.e., durability, crashworthiness, etc.). The system characteristics influence the performance in all the different attributes. Challenges arise when designing a system for improving mutually competing objectives, satisfying constraints from multiple engineering disciplines, and determining a single set of values for the vehicle characteristics. This presentation discusses an approach that conducts optimization analysis for a complex system by coordinating operations and exchange of data and information through a network of optimizations. A rotorcraft application is analyzed for structural-acoustic and crashworthiness. The hybrid finite element analysis (hybrid FEA) comprises the simulation method used for the structural-acoustic simulation. It combines conventional FEA with energy FEA for modeling the response of a system comprised by stiff, load bearing members, and flexible panels. The LS-DYNA solver is used for the crash-landing simulations. The weight of the gear box foundation is optimized while simultaneous improvements are pursued through parallel driven optimizations that address the input power into the system from gearbox excitation and the dynamic stresses encountered in the gearbox foundation from crash landing.

4:00

2pSA8. SALINAS: A massively parallel finite element code for structural dynamics and acoustics analysis. Jerry W. Rouse (Dept. Analytical Structural Dynam., Sandia Natl. Labs., Albuquerque, NM 87185, jwrouse@sandia.gov), Timothy F. Walsh, and Garth M. Reese (Sandia Natl. Labs., Albuquerque, NM 87185)

This talk shall present an overview of SALINAS, a massively parallel finite element code for structural dynamics and acoustics analysis that is being developed at Sandia National Laboratories. SALINAS allows for prediction of both the time and frequency domain responses of complex structural, acoustic, and fully coupled structural acoustic systems having millions of degrees of freedom. An overview of SALINAS capabilities shall be presented including development history, solver and element types, quadratic eigenanalysis and frequency response, direct frequency response, nonlinear acoustics, implicit transient dynamic analysis, and infinite elements with focus given to structural acoustics capabilities. The application of SALINAS to structural acoustics problems shall also be presented as well as future directions of research for the development of the code.

4:15

2pSA9. Numerical study of structural intensity for a model airframe structure in healthy and damaged states. Micah R. Shepherd, Stephen C. Conlon (Appl. Res. Lab., Penn. State Univ., P.O. Box 30, State College, PA 16804), and Fabio Semperlotti (Penn. State Univ., State College, PA 16804)

Structural damage due to fatigue or material flaws presents a major challenge for many engineering applications. To address this challenge, structural health monitoring techniques have been developed in an effort to determine the type, severity, and location of the damage. Many of these techniques are based on the nonlinear contact phenomenon. In this paper, structural intensity methods are used as a damage-detection metric on a simplified airframe structure. A finite-element model was created and the structural intensity field was mapped over the surface using individual element power flow estimates. Damage was then introduced into the model by removing connectivity between elements and including nonlinear stress-strain relationships and contact at the damage location. Structural intensity maps were then compared to the healthy case to correlate the changes with the damage state. More specifically, the increase in structural intensity at the sub- and super-harmonics was found to be indicative of the damage state. The results of this numerical study show the potential of surface structural intensity in autonomous damage detection for flight vehicles. The implications of this type of damage detection metric with respect to damage size and location will also be discussed.

Session 2pSC

Speech Communication: Speech Production (Poster Session)

Benjamin R. Munson, Chair
Univ. of Minnesota, Dept. of Speech, Language, Hearing Science, 164 Pillsbury Ave., SE, Minneapolis, MN 55455

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbers papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pSC1. A procedure for estimating gestural scores from articulatory data. Hosung Nam (Haskins Labs., New Haven, CT 06511, nam@haskins.yale.edu), Vikramjit Mitra (Univ. of Maryland, College Park, MD), Mark K. Tiede (Haskins Labs., New Haven, CT), Elliot Saltzman (Boston Univ., Boston, MA), Louis Goldstein (Univ. of Southern California, Los Angeles, CA), Carol Espy-Wilson (Univ. of Maryland, College Park, MD), and Mark Hasegawa-Johnson (Univ. of Illinois at Urbana-Champaign, Urbana, IL)

Speech can be represented as a set of discrete vocal tract constriction gestures (gestural score) defined at functionally distinct speech organs [tract variables (TVs)]. Using such gestures as sub-word units in an ASR system, variation in speech arising from coarticulation and reduction can be addressed. Since there is a lack of test corpora annotated with gestural scores, we develop a semi-automatic procedure for estimating and annotating gestural scores from natural speech databases using the Haskins speech production model (TaDA). We first describe the procedure's application to 500 words with unique phone sets found in the Wisconsin x-ray microbeam database, generating both gestural scores and synthetic speech. Second, we perform dynamic time warping (DTW) to align the TaDA-generated speech signals with respect to the microbeam data. The DTW time-scaling pattern is then used to adjust the gestural score originally input to TaDA to generate new time-warped TaDA acoustics. Third, we fine-tune the gestural timing patterns within the new gestural score in a constrained manner to generate the score whose associated acoustics are minimally different from the database acoustics. Finally, we evaluate the optimized gestural score's validity by comparing its associated TV trajectories to those from the original microbeam database.

2pSC2. Prosodic interaction between speakers of American and British English. Jelena Krivokapic (Dept. of Linguist., Yale Univ., 370 Temple St., 302, New Haven, CT 06520-8366, jelena.krivokapic@yale.edu)

Previous work has found that interaction between speakers influences their production, in that phonetic similarity of the speakers to each other increases [Sancier and Fowler, J. Phonetics **25**, 421–426 (1995); Pardo, J. Acoust. Soc. Am. **119**, 2382–2393 (2006); and Tobin, J. Acoust. Soc. Am. **125**, 2757 (2009)]. These studies have mostly focused on segmental properties of speech. Little is known, however, about how prosodic properties of speech are affected in conversational interaction. An experiment is presented that examines the interaction between native speakers of British English and American English. Using the synchronous speech paradigm [Cummins, ARLO **3**, 7–11 (2002) and Zvonik and Cummins, in Proceedings of the Eurospeech 2003 (2003), pp. 777–780] where two speakers read sentences at the same time, we examine how speakers influence each other at the prosodic level. Eight subjects (4 dyads), each consisting of one British and one American speaker, read the following: a short story that contained four words where the two dialects differ in stress pattern, 13 target words where the dialects differ in vowels, and 84 sentences with varying intonation patterns (question, statement, and focus). The goal was to examine how speakers adjust to each other's speech, and whether certain prosodic properties are more susceptible to adjustment to the co-speaker's production than others.

2pSC3. Prosodic boundaries in German: Final lengthening in spontaneous speech. Susanne Fuchs (ZAS, Schuetzenstrasse 18, 10117 Berlin, Germany fuchs@zas.gwz-berlin.de), Jelena Krivokapic (Yale Univ., New Haven, CT 06520-8366), and Stefanie Jannedy (ZAS, Schuetzenstrasse 18, 10117 Berlin, Germany)

Most theories of prosodic structure postulate at least two phrasal categories above the word level, a minor and a major one. One correlate of phrasal boundary marking is lengthening on the right edge of a phrase. To gain a theory neutral understanding of the nature of prosodic boundaries, a Gaussian mixed model (GMM) was applied to durational data, estimating the underlying clusters of a continuous distribution. Spontaneous speech data were collected with standardized interviews from 19 adolescent speakers of a multi-ethnolect (Kiezdeutsch, Hood German) spoken in Berlin, orthographically transcribed and added to a database that allowed for searches of the particle "so"(so or like) in different conditions. Acoustic durations of /z/ and /o/ of so in phrase final position (as determined by orthographical markings or by following pauses) were labeled and z-transformed per speaker. The

results show that a model including two clusters accounts best for the data. Nevertheless, there is an overlap between the two distributions supporting the gradient nature of boundaries. Further analyses with more tokens of so including word level junctures are currently under way. The impact of our data-driven results on theoretical models of prosodic boundaries is discussed. [Supported by a grant from BMBF.]

2pSC4. Prosodic analysis of subglottal pressure contours. Helen M Hanson (Dept. ECE, Union College, 807 Union St., Schenectady, NY 12308, helen.hanson@alum.mit.edu), Stefanie Shattuck-Hufnagel (MIT, Cambridge, MA 02139), and Angela McLelland (Union College, Schenectady, NY 12308)

In previous work the problems of identifying final fall and estimating declination of the subglottal pressure contour were addressed [Hanson *et al.*, J. Acoust. Soc. Am. **118**, 2027 (2005); **120**, 3090 (2006)]. Additional methods for identifying final fall have since been explored. These new methods will be described, and the identification of final fall will be compared across several of the methods. The use of these additional methods has meant that the complete data set for three subjects can be analyzed. Therefore it is worth revisiting the work presented in 2006. Objective measures of subglottal pressure declination will be evaluated with respect to subjective ratings of declination. The effects of phrase and boundary tones on declination rate will be re-examined, as will the effects of pitch accent distribution. [Work supported by NIH Grants DC00075 and DC04331 and by a Union College summer research grant to A.M.]

2pSC5. Effects of metrical structure on syllable timing II. Megan Tinius and Marios Fourakis (Dept. of Communicative Disord., UW-Madison, Madison, WI 53706)

Fourakis and Monahan [J. Linguist. (1988)] examined the effects of two different types of metrical foot (iamb and anapest) on syllable timing in 3 ft utterances. These utterances were comprised of three iambs, three anapests, and every possible combination of the two types. The present research follows the same paradigm but uses trochees and dactyls instead of iambs and anapests. A set of eight 3 ft long sentences was constructed. At one end there was the three trochee sentence: Auctions rescue business. At the other end was the three dactyl sentence: Auctions may rescue his businesses. The intermediate six sentences were formed by different combinations of the metrical feet. Five male and five female native speakers of Midwestern American English were recorded producing five repetitions of each sentence at a normal rate of speech. Segments were assigned syllable membership based on the maximum onset principle. Results will be presented on syllable durations as a function of foot type and sentence type. [Research supported by the University of Wisconsin Madison Graduate School.]

2pSC6. Temporal structure in the speech of persons with dementia: A preliminary study. Linda Carozza (Dept. of Commun. Sci. and Disord., St. John's Univ., 300 Howard Ave., Staten Island, NY 10301), Margaret Quinn, Jacqueline Palladino, and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

Although cognitive and language processes in dementia have been studied extensively, the question of motor speech degeneration in the course of dementing illness is a relatively unexplored area. The potential for early dissociation of motor functions of language at the level of speech production has not been explored; an interaction between motor speech and language production and perception changes should inform our understanding of the deterioration in dementia. If early motor declines are found, they may provide an additional early marker and contribute to early detection, giving further insight to the character of neurological deterioration in this disease. This study explores the temporal structure of speech in individuals in the mild-to-moderate stages of the disease. In this preliminary study, participants were asked to produce a series of four-word phrases containing target word in phrase-medial or phrase final position. The target words contained the vowel in "heed," "head," or "had," and began with a fricative or voiced or voiceless stop consonant and ended with either /t/ or /d/. Thus, we are able to examine four aspects of the temporal patterns of speech in these persons: intrinsic vowel duration, VOT, phrase-final lengthening, and the effect of final consonant voicing on vowel duration.

2pSC7. Japanese language learners' production of short/long vowel and single/geminate consonant contrasts: A longitudinal development. Yukari Hirata and Ian C. McNally (Dept. of East Asian Lang. and Lit, Colgate Univ., 13 Oak Dr., Hamilton, NY 13346, yhirata@colgate.edu)

This longitudinal study examined the development of native English speakers' production ability to distinguish between single/geminate consonant and short/long vowel pairs in Japanese over 4 months abroad in Japan. Seven participants, who had formal Japanese language classes for 2 years in the US, recorded rika(a) and ka(k)ko pairs in a carrier sentence three times each, at three speaking rates, before and after the study abroad (pre-/post-Japan). Seven measures, including the duration of contrasting segments, words, and sentences, were taken and various ratios were also calculated. Of the native Japanese speaker criteria used to evaluate participants' data, the vowel-to-word and closure-to-word duration ratios of 0.38 and 0.35 were used as boundary ratios between short and long segments [Hirata (2004); Hirata and Whiton (2005)]. Using these criteria, participants' data were classified at 66% accuracy at pre-Japan but improved to 85% at post-Japan for vowel distinction. In contrast, they improved from 73% to 74% accuracy for consonant distinction. Analyses of other durational measures also supported that participants made significant improvements in vowel distinction, but not in consonant distinction. These results support the claims of Toda (1997) and Han (1992) and extend to include intermediate language learners into their respective studies.

2pSC8. Modifying speech to children based on perceived developmental level: An acoustic study of adults' fricatives. Hannah M. Julien, Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., Minneapolis, MN 55455 julie006@umn.edu), Jan Edwards (Univ. of Wisconsin, Madison, WI 53705), Mary E. Beckman, and Jeffrey J. Holliday (Ohio State Univ., Columbus, OH 43210)

Recent acoustic studies have shown that children learning English gradually differentiate between the sounds /s/ and /ʃ/ [Li *et al.* (2009)]. Other studies have shown that adults can perceive different degrees of accuracy in these phonemes when given a non-categorical response modality, such as a visual-analog scale [Urberg-Carlson *et al.* (2008)]. This presentation combines a VAS rating task with a production task. Adults were presented with children's productions of target fricatives. They rated the accuracy, after which they were instructed to produce the target sequence as if they were responding to the child whose speech they had just rated. Ongoing analyses examine whether adults respond differently to productions that they have rated as intermediate between /s/ and /ʃ/ and whether adults produce hyperarticulated version of fricatives in response to productions that they rated as not canonically /s/ or /ʃ/ , as opposed to ones that they rated as adult-like. The results of this project will be used in supervised learning models of the acquisition of /s/ and /ʃ/ and will inform us on the extent to which adults calibrate their speech to children depending on the perceived level of speech mastery. [Work supported by NIDCD Grant 02932 and NSF Grant BCS0729277.]

2pSC9. Palatalization in infant-directed speech. Jenesia McCammon and Nan Bernstein Ratner (Dept. of Hearing and Speech Sci., Univ. of Maryland, 0100 Lefrak, College Park, MD 20742, jmccammon518@gmail.com)

Adult-directed speech (ADS) is characterized by frequent use of phonological rules such as palatalization ("did you" becomes "didju"). Use of such rules should blur word boundaries, making it more difficult for infants to identify words in the input. To date, there are conflicting data on the degree to which infant-directed speech (IDS) makes more or less frequent use of these rules [Shockey and Bond (1980) and Bernstein Ratner (1986)]. We are following a cohort of children from 7 months to 2 years. We examined whether rules such as palatalization are used more or less frequently in IDS by 20 mothers to their children and to an adult listener. To date, there are more opportunities for palatalization in the mothers' IDS than ADS to preverbal children, and the rule is more frequently applied in IDS. In this case, IDS is not more acoustically clarified than ADS. Research suggests that IDS

clarification may be a function of the child's language stage. We would therefore not expect reduction in palatalization, which primarily affects functors, until a later stage of the child's language development.

2pSC10. Voice onset time in infant-directed speech at 7.5 and 11 months. Anna Synnestvedt, Nan Bernstein Ratner, and Rochelle Newman (Dept. of Hearing and Speech Sci., Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742)

Studies have reported differences between infant-directed speech (IDS) and adult-directed speech (ADS), suggesting that mothers adjust speech to their infants in ways that may help children better parse the incoming acoustical signal. One aspect of IDS that has been examined is voice onset time (VOT). Results have been inconsistent, revealing longer VOT in IDS [Englund (2005); Malsheen (1980)], shorter VOT in IDS [Sundberg & Lacerda (1999)], or no difference in VOT between ADS and IDS [Baran *et al.* (1977)]. Characteristics of IDS may also depend on the language maturity of the child and, therefore, clarity of speech may vary across stages of development [Englund & Behne (2006)] as well as vary among mothers. The present study examines 15 mothers' VOT in IDS to children at 7.5 months old and again at 11 months as compared to their VOT in ADS. Words with initial stop consonants that occurred in both IDS and ADS conditions are analyzed using PRAAT. Variability was observed among mothers and age effects: preliminary data suggest that VOT in IDS at 11 months is better differentiated than at 7.5 months. [Work supported by NSF BCS0745412.]

2pSC11. Mandarin and English initial stops produced by Mandarin-English bilinguals. Jia-Shiou Liao (School of Appl. Foreign Lang., Chung Shan Medical Univ., No. 100, Sec. 2, Daching St., South District, Taichung City 402, Taiwan, jsliao@csmu.edu.tw)

This study investigates the relationship between the way Mandarin-English bilinguals and native speakers of English pronounce English initial stop consonants, and between the way Mandarin-English bilinguals pronounce Mandarin and English stops, with respect to voice onset time. Twelve Mandarin subjects who had learned English as their second language and were studying at a US university were asked to read randomized lists of 72 Mandarin characters and of 72 English words; twelve English subjects read only the list of English words. The syllable structure of each Mandarin character and of each English word in this study was CV. The Mandarin consonants were initial aspirated and unaspirated /p, t, k/; the English consonants were initial /p, t, k/ and /b, d, g/. The vowels used in both lists were high vowels. Significant differences are found between the Mandarin-speaking subjects' articulation of all the Mandarin stops and their minimally paired English equivalents produced by the English-speaking subjects. Significant differences are found between Mandarin and English speakers' articulation of the same English stops, except for that of /t/. When comparing the Mandarin-speakers' articulation of Mandarin stops and their minimally paired English counterparts, a significant difference is found only between English and Mandarin aspirated /p/.

2pSC12. Phonetic adaptation in non-native speech: Insights from a distributional analysis of long-lag voice onset time. Marisa Pineda and Meghan Sumner (Dept. of Linguist., Stanford Univ., Margaret Jacks Hall, Bldg. 460, Stanford, CA 94305-2150, middyp@stanford.edu)

Speakers encountering long-lag voice onset time (VOT) for the first time in their L2 produce VOTs between their L1 and L2 values. Native-like long-lag productions are conditioned by speaker competency factors such as age of acquisition and experience, showing significant production differences between late bilinguals, early bilinguals, and native L2 speakers. Thus far, analyses have focused on average VOTs across speaker groups. This project investigates the full distributional properties of VOT in bilinguals (e.g., variation and skewness) in addition to averages, providing a more informative picture of bilingual acquisition. VOT production data were collected from French-English bilinguals (age of English onset 0–15 years) and submitted to a distribution-based analysis. Results show that while speaker groups differ predictably in mean VOT, the analysis discovers subgroups based on common production behaviors, eliminating the gross categorizations of early and late bilinguals and moving toward gradual, predictable shapes of VOT that are correlated with English schooling, time in an English-speaking country, and age of acquisition. Ultimately, these results may fill in some missing blanks between perception and production, suggesting, for example, that differences in perceived accent and comprehensibility diverge due to degree of overlap with native L2 VOTs.

2pSC13. A Bayesian model of voice-onset time production. Mark VanDam (Boys Town Natl. Res. Hosp., 555 N 30 St., Omaha, NE 68131, mark.vandam@boystown.org) and Noah Silbert (Indiana Univ., Bloomington, IN 47405)

Talkers performed a listen-and-repeat task to investigate temporal detail in voice-onset time (VOT) productions of American English word-initial stop consonants. Experimental factors included linguistic context (isolation, carrier phrase, unfamiliar phrase, and familiar phrase), usage frequency (high and low), lexical status (word and non-word), training (baseline and posttraining), and posttraining generalization (test words and novel words). For each context, frequency, and lexical status, baseline VOT production estimates were collected, then a naive training regimen conducted, then posttraining estimates of both test words and novel words were obtained. Testing novel words explored whether the effect, if obtained, generalized throughout the lexicon. A Bayesian linear model (analogous to analysis of variance) was used to model VOT means as a function of these factors. Posterior distributions of modeled VOT means were compared across six talkers, with a focus on probing the relationships between lexical frequency and status, linguistic context, and training. Preliminary results suggest that a number of these experimental factors influence fine-grained control of VOT production. As expected from speech production data obtained from multiple talkers and thousands of productions, the overall model fit was imperfect, but the results indicate that a Bayesian model can be productively deployed for data exploration into temporal aspects of speech production.

2pSC14. Priming at the level of phonetic detail: Evidence from voice onset time (VOT). Susannah Levi and Jennifer Bruno (Dept. of Communicative Sci. and Dis., NYU, 665 Broadway, 9th floor, New York City, NY 10012, svlevi@nyu.edu)

Priming studies have shown effects on the production of target words when the target and prime are identical and when the target and prime share some but not all phonological features. Variation at the level of phonetic detail has also been found to affect production in word shadowing tasks where speakers produce longer VOTs when the VOT of the sample word is artificially lengthened. In this study, we examine the effect of naturally produced, yet systematic within-category variability in the VOT of the prime on the production of a target word for visually presented words. Primes either contained a voiceless stop with a long VOT ("keen") or a short VOT ("pan"). The target words contained a voiceless alveolar stop followed by five vowels which varied in height and backness (e.g., "tune" and "ten"). Results showed that the same target word was produced with longer VOTs when the prime contained a naturally long VOT than when the prime contained a naturally short VOT. These results suggest that speakers are sensitive to natural within-category variability, and that this sensitivity affects production across different segments.

2pSC15. Why are Korean tense stops mastered early: Evidence from production and perception. Eun Jong Kong (Waisman Ctr., Univ. of Wisconsin at Madison, 1500 Highland Ave. #493, Madison, WI 53705, ekong@wisc.edu) and Mary E. Beckman (The Ohio State Univ., Columbus, OH 43210)

Transcription-based studies have found that Korean-acquiring children master tense stops earliest among the three different types of homorganic stops (tense vs lax vs aspirated) despite its phonologically marked status. Tense stops in Korean have a short-lag VOT, so this finding is consistent with previous cross-linguistic research on order of acquisition of stop phonation types. However, Korean tense, lax, and aspirated stops are also differentiated by the fundamental frequency and the voice quality at the vocalic onset in addition to VOT. This study examined how correctly these multiple acoustic cues (VOT, f0, and H1-H2) predict the mastery pattern of Korean stops. The effect of these acoustic parameters on 20 Korean adult listeners' assessments of children's (aged 2–6) productions of /t′/, /t/, and /th/ was analyzed. Listeners were asked to label the stimuli as one of three stop cat-

egories and then to rate their goodness using visual analog scaling. Both categorical and gradient assessments of children's stops were analyzed based on the three acoustic parameters in a mixed-effect regression model in order to quantify the relative effects of acoustic parameters on adult perception of phonation. [Work supported by NIDCD 02932 and NSF Grant 0729140.]

2pSC16. "Probably, OK, whatever!": Variability in conversational speech stops and flaps. Natasha Warner (Dept. Linguist., Univ. of Arizona, Box 210028, Tucson, AZ 85721-0028, nwarner@u.arizona.edu) and Benjamin V. Tucker (Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

In casual, conversational speech, speakers often do not fully realize all segments one would normally expect in a given word. For example, "better" may be pronounced with an approximant instead of a medial flap (if the tongue fails to make a complete closure at the alveolar ridge), the expected flap may be entirely deleted, or the segments may be so heavily overlapped that it is almost impossible to say which segments are present and which not. The current work investigates how much reduction occurs (including both changes and deletions), comparing casual conversations, connected reading, and isolated word-list reading. It compares the phonemes /p, t, k, b, d, g/, post-stress vs inter-unstressed, and six segmental environments. The results represent 13 native English speakers and comprise a dataset of more than 9000 tokens, with several acoustic measures. This provides very thorough data on phonetic variability in both natural and "lab" speech. The work finds relatively little effect of stress and complex effects of segmental environment attributable to places of articulation. It also provides insight into how variable speakers are in how much they reduce, both within a speaker and within the population, a topic not often addressed in phonetic research.

2pSC17. Functional anatomy of the human tongue: Review and implications for the development of speech production. Margaret Denny and Richard S. McGowan (CReSS LLC, 1 Seaborn Pl. Lexington, MA 02420)

To acquire speech children must attempt to reproduce adult acoustic models using immature vocal tracts which differ from those of adults in complex ways. Control of the tongue is crucial for acceptable speech, yet the tongue is an understudied articulator because it is anatomically complex and difficult to record from using kinematic or electrophysiological methods. Nevertheless recent advances in tongue anatomy and physiology, combined with knowledge of vocal tract development, promise to shed light on the problems that children must solve in learning to speak as well as the solutions available to them. This review aims to integrate and present these findings to an audience interested in the development of speech production. A striking finding is that the tongue shows an anterior-to-posterior gradient in muscle fiber sizes and types. In multiple muscles, faster fibers are more common anteriorly while slower fibers are more predominant in the posterior tongue. Within fiber type (fast or slow) smaller fibers are seen anteriorly and larger ones posteriorly. This implies that rapid fine positioning and shaping may be particularly challenging when it involves the posterior tongue. [Work supported by NIDCD-0001247 to CReSS LLC.]

2pSC18. Tracking muscle deformation during speech from tagged and diffusion tensor magnetic resonance imaging. Xiaofeng Liu (Dept. of Comput. Sci., Johns Hopkins Univ., 1600 N. Charles St., Baltimore, MD 20017, xiaofeng.liu@gmail.com), Sudarshan Ramenahalli (Johns Hopkins Univ., Baltimore, MD 20017), Hideo Shinagawa, Maureen Stone (Univ. of Maryland Dental School, Baltimore, MD 21210), Jerry L. Prince (Johns Hopkins Univ., Baltimore, MD 20017), Emi Z. Murano (Johns Hopkins Hospital, Baltimore, MD 21287), Jiachen Zhuo, and Rao Gullapalli (Univ. of Maryland Med. School, Baltimore, MD 21210)

Oral head and neck cancer is the sixth most common cancer worldwide. These tumors are usually treated by surgical removal of the affected tissue. The result of the surgery is a loss of muscle tissue, accompanied by scarring, reduced strength, and often reduced function. This paper explores the relationship between tongue muscle orientation and muscle deformation pattern in normal and post-cancer surgery speakers. Our motivation is the need to better understand the mechanisms that underlie tongue motion, in order to better interpret clinical observations and to provide data that can help predict optimal surgical outcomes. This study uses diffusion tensor imaging (DTI) to extract muscle fiber orientation, but it is not possible to image the muscle deformation in real-time using DTI. On the other hand, tagged MRI can track muscle fiber bundles in real time, but the original fiber bundle is not visible. Therefore, this work develops a method to overlay muscle bundles from the DTI onto the corresponding locations in tagged MR images and track muscle deformation. The mechanical properties of the muscles are calculated, such as translation and bending. Motion pattern differences between the normal and glossectomy tongue are detailed. [NIH R01-CA133015 and K99-DC9279.]

2pSC19. Dependency of compensatory strategies on the shape of the vocal tract during speech perturbed with an artificial palate. Jana Brunner (Res. Lab. of Electron., MIT, 50 Vassar St., Cambridge, MA 02139, jbrunner@mit.edu), Phil Hoole (Ludwig-Maximilians-Universität München, 80799 München, Germany), Frank H. Guenther (Boston Univ., Boston, MA 02215), and Joseph S. Perkell (MIT, Cambridge, MA 02139)

When adapting to an articulatory perturbation, different speakers may use different strategies. This study investigates the hypothesis that the choice of strategy depends on vocal tract shape. Vowel production was perturbed with prostheses that changed palatal shape. Three speakers had a prosthesis that effectively moved the alveolar ridge posteriorily; three others had a prosthesis that effectively flattened the palate. EMA recordings of speakers' adaptive behavior showed that during the production of the vowel /y/, the speakers with the alveolar prosthesis compensated by using increased lip protrusion. The speakers with the flat prosthesis used about the same amount of lip protrusion as without perturbation. We hypothesize that this is because for the first group of speakers the main constriction in the vocal tract was moved forward due to the changed palatal outline: the front cavity became too short, which was compensated for by more lip protrusion. This hypothesis is being explored by determining whether the DIVA speech production model will produce compensations like the subjects when the model's vocal tract is adapted to fit individual vocal tract shapes using detailed information about the subjects' vocal tracts from MRI recordings. Complete results will be reported and discussed. [Work supported by DAAD, NIH.]

2pSC20. Can an ultrasound tongue be overlaid on an MRI (magnetic resonance imaging) vocal tract? Jim Lee (jima@yahoo.com) and Maureen Stone (Dept. of Neural and Pain Sci. and Dept. of Orthodontics, Univ. of Maryland Dental School, Baltimore, MD 21210)

Ultrasound image sequences and cine-MRI time-frames provide intersecting but not identical data about the articulation of speech. It would be valuable to be able to combine ultrasound and MRI datasets in a way that maximizes their strengths. Midsagittal MRI images the entire vocal tract; ultrasound images the tongue surface. However, MRI is not portable, requires multiple repetitions, and supine subject positioning. Thus insertion of the tongue from ultrasound movies onto the MRI vocal tract would provide tongue motion in a head coordinate system. This study aligns supine ultrasound tongue contours with MRI tongue contours of the same tasks using (1) rigid alignment of the tongue contours frame-by-frame and (2) rigid alignment of palates, whose alignment parameters are then applied to tongue contours. rms differences compare the error between methods. The value of this study is that if palatal alignment yields a comparable error to tongue alignment, then palatal alignment could be used to insert an ultrasound data set into a vocal tract outline that has no tongue motion data associated with it. [Funding support: R01CA133015.]

2pSC21. An articulatory, acoustic, and perceptual study of the effect of a short lingual frenulum on flaps in late Korean-English bilinguals. Kyung-Ah Kim (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853, kk339@cornell.edu)

Having a short lingual frenulum is known to limit tongue mobility, often resulting in nonstandard pronunciations of certain coronal sounds. While previous studies have largely focused on the physical and linguistic characteristics of the most extreme cases and on the effects of surgical intervention [Hagiwara *et al.* (2002); Lee *et al.* (2008)], the current study investigates the articulatory, acoustic, and perceptual characteristics associated with Korean speakers with a relatively short lingual frenulum but not short enough to

greatly impede communication. The study focuses on the flap—the single most problematic sound for Korean speakers with short lingual frenula—produced in both Korean and English by both "short-tongued" and normal late Korean-English bilinguals. Flaps are studied through palatographic, linguographic, and spectrographic analyses, while a perception test is performed to pinpoint the categorical boundary between the normal and short-tongued flaps. Preliminary results suggest that the short-tongued speakers are better able to accurately produce the English flap than the Korean flap, due to language-specific phonetic and phonological characteristics. This finding bears on whether surgical intervention is necessary to help normal (and mildly short-tongued) speakers of Korean improve their English pronunciation [se Ko (unpublished)].

2pSC22. Resonance-based source-filter separation for speech. Ivan Selesnick (Dept. of Elec. and Comput. Eng., Polytechnic Inst. of New York Univ., 6 Metrotech Ctr., Brooklyn, NY 11201) and Chin-Tuan Tan (New York Univ., New York, NY 10016)

This paper proposes the application of resonance-based signal analysis to the decomposition of speech into "source" and "filter" waveform components. Resonance-based signal analysis, developed by Selesnick,[1] is a novel analysis method that decomposes a signal into high-resonance (oscillatory) and low-resonance (transient) components. The original signal is given as a sum of the two components. The method utilizes sparse signal representations in the domain of two invertible constant-Q transforms, having low- and high-Q-factors, respectively. The constant-Q transforms are implemented as rational-dilation wavelet transforms. Applied to speech, the method provides a separation that cannot be achieved with other known filtering techniques. Current source-filter modeling of speech assumes a known linear filter which sufficiently estimates vocal tract characteristics and can be derived from a segment of stationary speech. The source glottal waveform is then extracted by inverse filtering. In contrast, resonance-based signal analysis directly decomposes speech into its source and filter waveforms without such assumptions. The outcome may provide a new perspective in speech analysis, synthesis, recognition, and coding. This paper illustrates the efficacy of this signal analysis technique with vowels; further investigation with other speech elements is in progress.

2pSC23. Speech separation based on multi-level segmentation. Jose J. Lopez and Maximo Cobos (ITeam, Univ. Politecnica Valencia, Camino de Vera, 46022 Valencia, Spain, jjlopez@dcom.upv.es)

Speech separation is still a challenging issue in acoustic signal processing. To deal with this problem, sparse methods are usually employed when there are more sources than sensors (underdetermined problem). In this paper, a two-microphone separation method is presented. The proposed algorithm is based on grouping time-frequency points with similar direction-of-arrival (DOA) using a multi-level thresholding approach. The thresholds are calculated via the maximization of the interclass variance between DOA estimates and allow to identify angular sections wherein the speakers are located with a strong likelihood. Separation is finally performed by means of time-frequency masks. Several experiments conducted under different mixing scenarios are discussed, showing the benefits of the proposed approach.

2pSC24. Classification of gender based on cepstral coefficients and spectral moments. Laura Spinu and Jason Lilley (Univ. of Delaware, Newark, DE 19716)

Several gender classification methods based on acoustic information were compared. The data came from 31 native speakers of Romanian (10 males, 21 females). A subset of fricatives and vowels (7348 tokens) was divided by hidden Markov model training into three acoustically uniform regions. For each region, (a) a set of cepstral coefficients, specifically c0–c4 and (b) two sets of spectral moments, specifically bark-transformed and linear moments 1–4 plus rms, were extracted. The acoustic data were then used in a linear discriminant analysis to classify the tokens by gender. The findings show that cepstral coefficients perform better than spectral moments in gender classification, and that spectral moments are more successful than linear moments. The overall correct classification was 90% for the cepstral analysis, 78% for bark spectral moments, and 72% for linear moments. When the data from fricatives and vowels were examined separately, it was found that in all cases the vowel information yielded more accurate gender classification (e.g., 84% correct classification for cepstral moments based on vowels only), but the fricative information alone could account for 78% correct classification. These analyses provide insight into the differential distribution of acoustic features related to gender over segment types.

2pSC25. Manipulating the short time Fourier transform and cross-spectrum to estimate speech parameters. Douglas J. Nelson (Natl. Security Agency, 9800 Savage Rd., Fort Meade, MD 20755)

Presented are methods for automated estimation of speech parameters, including pitch, formant frequencies, and the instants of excitation of voiced speech. These methods are based on reassignment applied to the short time Fourier transform (STFT) and to the cross-spectrum computed as the product of the STFT and the complex conjugate of the STFT delayed in time. The cross-spectrum is a complex-valued surface whose magnitude is the conventional spectrogram and whose phase is the phase of the STFT differentiated with respect to time. It is demonstrated that this representation may be easily manipulated to detect and extract information, such as pitch and formant frequencies. Based on the cross-spectral representation, we present a peak representation that may be used to efficiently encode speech spectral features, and we demonstrate applications of this representation.

Session 2pUW

Underwater Acoustics, Acoustical Oceanography, and Signal Processing in Acoustics: Acoustic Particle Velocity and Vector Fields: Signal Processing and Communication Applications II

Mohsen Badiey, Cochair
Univ. of Delaware, College of Earth, Ocean, and Environment, Newark, DE 19716-3501

Ali Abdi, Cochair
New Jersey Inst. of Technology, 323 King Blvd., Newark, NJ 07102

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. Modeling the acoustic channel for mixed pressure and particle velocity sensors. Paul Hursky and Michael B. Porter (HLS Res. Inc., 3366 North Torrey Pines Court, Ste. 310, La Jolla, CA 92037)

Looking at our repertoire of acoustic modeling techniques, ray tracing and Gaussian beams provide a good match to modeling the underwater channel for acoustic communications, given their fundamentally broadband nature. We will discuss how we have augmented existing acoustic models, including Gaussian beam and normal mode codes, to produce particle velocity predictions. In both of these cases, the modifications consisted of manipulating the components used to represent the pressure field, as opposed to relying upon a finite difference approximation of the predicted field itself. We will discuss further modifications to these codes that enable the dynamic features of the channel to be modeled, including source and receiver motion, and motion of the ocean surface due to ocean surface waves. It is important to model realizations of the channel, rather than averaged representations, because many of the high-speed, coherent modulation schemes that are the focus of continuing research use tracking loops in their design. These tracking loops must be stressed with realistically time-evolving channels. Acoustic channel predictions will be compared with experiment data from acoustic communications experiments off the coast of Kauai in Hawaii.

1:25

2pUW2. Geo-acoustic inversion with vector sensors in shallow water. Fenghua Li and Renhe Zhang (Natl. Lab. of Acoust., Inst. of Acoust., Chinese Acad. of Sci., Beijing 100190, China, lfh@mail.ioa.ac.cn)

Vector sensors have attracted much attention in recent years. However, relative few papers have been published on the geo-acoustic inversion with vector sensors. The data recorded with vector sensors in several shallow water acoustic experiments have been studied in this paper. The experimental data and theoretical analyses indicate that (1) the intensity of the vertical particle velocity decreases faster than that of the horizontal particle velocity; (2) the same normal modes of the horizontal particle velocity and the vertical particle velocity have similar transmission losses, but different amplitudes and phases; (3) the difference of the mode amplitudes between the horizontal particle velocity and the vertical particle velocity is dependent on the eigen-value and receiver depth. With those properties, two geo-acoustic inversion schemes employing vector sensors have been developed. The first inversion scheme uses a combination of matched field processing and the difference of transmission losses between pressure and particle velocity. The second estimates the bottom sound speed and attenuation from the amplitudes of normal modes of particle velocities. Experimental results show that the developed geo-acoustic inversion methods with vector sensors can decrease the uncertainty of inversion in comparison with that by hydrophones.

Contributed Papers

1:45

2pUW3. Modulated helicity for acoustic communications and helicity-selective acoustic receivers. Timothy M. Marston and Philip L. Marston (Dept. Phys. and Astronomy, Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu)

Helicoidal or vortex beams have an azimuthal phase gradient and an axial null in amplitude. The sign and magnitude of the azimuthal phase gradient determine the helicity and topological charge of the beam and the associated handedness. Beams with a unit-magnitude charge have been generated with a four-sector array, adjacent quadrants being driven with a 90 deg phase offset [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **106**, 3313–3316 (1999)]. The present work shows that the helicity can be rapidly modulated with an appropriately timed reversal of the excitation to a single pair of diagonal sectors. This reversal was demonstrated with an electronic commutator that inverts (at a zero crossing of the sine-wave excitation) two of the channels prior to amplification. While the wave helicity may be inferred with processed signals from a detection array, it may also detected with simple superposition of outputs from a four-element detector using appropriately offset elements. Digital communications was demonstrated where signals of the wrong helicity were suppressed. Scattering by symmetric objects on the axis of the beam preserves the helicity of the radiation [P. L. Marston, J. Acoust. Soc. Am. **124**, 2905–2910 (2008)]. [Work partially supported by ONR.]

2:00

2pUW4. Amplitude, phase, location, and orientation calibration of an acoustic vector sensor array, Part I: Theory. Hans-Elias de Bree (Microflown Technologies, P.O. Box 300, 6900 AH Zevenaar, debree @microflown.com)

An acoustic vector sensor array consists of multiple sound pressure microphones and particle velocity sensors. A pressure microphone usually has an omni-directional response, yet a particle velocity sensor is directional. Currently, acoustic vector sensor arrays are under investigation for far field source localization and visualization. One of the major practical issues in these applications, however, is to determine the accurate position, orientation, and complex (phase and amplitude) sensitivity of each sensor within the array. In this study, a calibration method is developed to determine each of those crucial parameters based on a limited number of measurements with a reference sensor and multiple sound sources located at known locations. The calibration method is also designed to be robust to mistakenly switched cable connections. Ideally, the calibration process should take place in an anechoic environment, but efforts have been made to compensate the effects of moderate background noise and reflections.

2:15

2pUW5. Amplitude, phase, location, and orientation calibration of an acoustic vector sensor array. Part II. Experiments. Tom Basten (Microflown Technologies B.V., Ruitenberglaan 26, 6800AH, The Netherlands, basten@microflown.com)

An acoustic vector sensor array consists of multiple sound pressure microphones and particle velocity sensors. A pressure microphone usually has an omni-directional response, yet a particle velocity sensor is directional and usually has a response pattern as a figure of number eight. Currently, acoustic vector sensor arrays are under investigation for far field source localization and visualization. One of the major practical issues in these applications, however, is to determine the accurate position, orientation, and complex (phase and amplitude) sensitivity of each sensor within the array. In this study, a calibration method, which is discussed in Part I, is verified with experiments. The method determines all the crucial parameters based on a limited number of measurements with a reference sensor and multiple sound sources located at known locations. The calibration method shows also to be robust to mistakenly switched cable connections. The experiments are performed in different acoustic environments including an anechoic room. The results will be presented in this paper.

2:30

2pUW6. Experimental study of vector array time reversal communication. Zhongkang Wang, Hongtao Zhang, and Xiaofan Yang (Hangzhou Appl. Acoust. Res. Inst., 96 Huaxing Rd., Hangzhou City, P.O. Box 1249, China)

The cooperative array performance experiment (CAPEx) was performed in Lake Washington near Seattle in September 2009 [Rouseff *et al.*, this session]. The experiment included a passive time-reversal communications component. A variety of communications signals in the 1.5–4 kHz band were transmitted including BPSK, QPSK, DSSS, and OFDM. The signals were recorded on an eight-element vertical vector array used as the receiving array at ranges up to 4 km. Results are given by using combined scalar and vector sensors. [Experiment supported by ONR.]

2:45

2pUW7. Proof of principle for geo-acoustic inversion of vector sensor array data. Robert A. Koch (Appl. Res. Labs., Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758, koch@arlut.utexas.edu)

Acoustic data collected from an array with elements comprised of omnidirectional hydrophone sensors co-located with acceleration vector sensor triplets are analyzed for seabed geo-acoustic information. The data were collected during the passage of a surface vessel, the R/V Montague, in an August 2006 experiment conducted in Monterey Bay. Geo-acoustic inversion with the vector sensor acceleration components parallel to the array line of bearing produced a solution approximately identical to the solution obtained from the inversion of the hydrophone array data. [Sponsored by ONR.]

3:00—3:30 Break

3:30

2pUW8. Concerning the null contours of vector sensors. Dean J. Schmidlin (2629 US 70 E, Apt. E2, Valdese, NC 28690-9005)

A common use of a vector sensor is to maximize the sensing of a plane wave coming from one direction while creating a null in regard to a plane wave coming from another direction. It is shown that when this is accomplished, there is a null not only in the desired direction but a whole contour of nulls. The shape of the contour is a function of a single parameter and is described by means of a contour plot in the azimuthal angle–elevation angle plane. Additional insight is found by mapping the contours onto a unit-sphere in the spatial domain. One byproduct of the notion of null contours is a method for selecting the coefficients of the vector sensor for the case of two interfering sources. Finally, it is shown that once the null contour is understood for the vector sensor, it is a simple matter to extend the concept to directional acoustic sensors of higher order.

3:45

2pUW9. Utilization of the pressure gradient along a sparse vertical line array to determine vertical arrival angle from partial reflections from a layered seabed. David R. Dall'Osto and Peter H. Dahl (Dept. of Mech. Eng and Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Pressure signals measured by two four-element vertical line arrays, centered at depths 25 and 50 m, during an experiment off the New Jersey coast in Summer 2006 (SW06, depth 80 m) are combined to determine the pressure gradient of the received signal. The pressure gradient is converted to an estimate of vertical particle velocity, which combined with an estimate of the mean-squared pressure, and the plane wave impedance establishes a non-dimensional quantity. This quantity is inverted to determine arrival angle estimates from separate multipath arrivals of the transmitted signals (1 kHz). A geoacoustic model for the seabed [Choi *et al.*, J. Acoust. Soc. Am. **124** EL128–EL134 (2008)] includes a prominent sub-bottom reflector at depth 20 m below water sediment interface, which produces an energetic arrival time separated from the bottom arrival. The arrival angle estimates are examined as the source moves out in range from 50 to 300 m, and results are compared to arrival angle estimates derived from PE simulations. The performance of this gradient-based estimator is evaluated for various types of noise sources and levels, utilizing both real and PE-simulated time series. [Work supported by ONR.]

4:00

2pUW10. A characterization of scatterred acoustic intesity fields in the resonance region. Robert J. Barton, III (Sensors & Sonar, NUWC Code 1522, 1176 Howell St., Newport, RI 02841, robert.barton@navy.mil) and Kevin B. Smith (Naval Postgrad. School, Monterey CA, 93943, kbsmith @nps.edu)

In this study, we investigate the properties of the scattered acoustic vector fields generated by simple rigid and fluid-filled spheroids. Analytical solutions are derived from acoustic target strength scattering models in the near field region. Of particular interest is the understanding of the characteristics of energy flow of the scattered acoustic vector field in the near- to far-field transition region. We utilize the separable active and reactive acoustic intensity fields to investigate the structural features of the scattered field components. Numerical and in-air measured results are presented for the near and transition region scattered acoustic vector field of simple spheroids. A qualitative method is developed for characterizing submerged scatterer properties based on boundary condition and features of the scattered vector field.

4:15

2pUW11. Three dimensional sound source localization and sound mapping using a p-u sensor array. Jelmer Wind (Microflown Technologies B.V., 6800AH Zevenaar, The Netherlands, wind@microflown.com)

An acoustic sensor array, which consists of pressure and particle velocity sensors, is an attractive alternative to phased pressure array because knowledge of the three dimensional (3-D) particle velocity directly characterizes

the direction of the source. Hence, it is possible to localize sources in the low-frequency range, where the phase difference between pressure sensors is small, and in the high-frequency range, where aliasing occurs with pressure sensors. This article applies advanced source localization techniques from aeroacoustics to acoustic vector sensors. Several methods are simulated and tested with several configurations and validated using measurement data from an anechoic chamber. The setup to test the algorithms consists of 24 sources that are distributed in 3-D around the set ups in the anechoic chamber. The sources are uncorrelated driven with white noise. Conventional beamforming methods have been compared with acoustic vector based deconvolution methods and MUSIC algorithms. Four configurations are tested: (1) a planar array of sound pressure microphones, (2) a planar array of pu probes (pu probes are collocated sound pressure and particle velocity probes); the velocity probes where the velocity probes are oriented in randomly, (3) a velocity gradient array, (4) a quad array of three dimensional velocity probes plus collocated sound pressure. The different methods and algorithms

TUESDAY EVENING, 20 APRIL 2010 8:00 P.M.

OPEN MEETINGS OF ASA TECHNICAL COMMITTEES

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held immediately after the Social Hours. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings, including proposals for special sessions, workshops, and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

Committees meeting on Tuesday are as follows:

Acoustical Oceanography	Laurel C/D
Engineering Acoustics	Laurel A/B
Musical Acoustics	Galena
Noise	Grand Ballroom III/IV
Physical Acoustics	Grand Ballroom VII/VIII
Psychological and Physiological Acoustics	Grand Ballroom I/II
Structural Acoustics and Vibration	Dover A/B

Session 3aAA

Architectural Acoustics and Noise: Primary and Secondary School Special Function Spaces I

Robert C. Coffeen, Chair
Univ. of Kansas, School of Architecture and Urban Design, Marvin Hall, Lawrence, KS 66045

Chair's Introduction—8:00

Invited Papers

8:05

3aAA1. The rest of the school: An investigation into good acoustical design of non-classroom spaces in primary and secondary schools. Benjamin Markham and Jennifer Hinckley (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

The focus on acoustics in recent standards and guidelines such as the ANSI Standard S12.60, the Leadership in Energy and Environmental Design (LEED) for Schools system, and the Collaborative for High Performance Schools (CHPS) has rightly been on classrooms. Each of them does address other school environments also, although sometimes without particular specificity. After reviewing relevant American standards and guidelines for gymnasiums, libraries, lobbies, corridors, auditoriums, and other spaces that do not fit neatly into the classroom designation, this paper addresses good practices in the acoustical design of such critical environments through the use of a series of case studies. Specifically, the authors suggest room acoustics design parameters and appropriate background noise and sound isolation criteria for non-classroom spaces in primary and secondary schools.

8:25

3aAA2. Case study in cafeteria noise solutions. Kenneth Good and Sean Browne (Armstrong World Industries, 2500 Coulumbia Ave., Lancaster, PA 17601)

A recent case study was conducted addressing excessive occupant noise within a cafeteria in a Rhode Island school. A design requirement was to retain the open structure ceiling. This paper will discuss the before and after acoustic performance and the effect it had on the occupants.

8:45

3aAA3. The acoustical balancing act for music rehearsal spaces in schools. David L. Adams (D. L. Adams Assoc., Inc., 1701 Boulder St., Denver, CO 80211, dadams@dlaa.com)

Conflicting acoustical criteria are encountered in the design of music rehearsal spaces, particularly when a single space must serve both orchestra and band rehearsals, not to mention the situation when the space must also serve for chorus rehearsals. For example, the criterion commonly found in the literature for room volume or more explicitly ceiling height conflicts with the requirement for an appropriate reverberation time or design goal loudness level, which are also dependent on the group (band, orchestra, and chorus) that will use the space and for that matter on the individual music teacher. The degree of involvement of music teachers in the design of these spaces and their potential impact on the end result is discussed. Case histories are presented to illustrate the good, the bad, and the ugly with respect to music rehearsal spaces in schools.

9:05

3aAA4. Acoustical considerations for vocal music rehearsal rooms: A choir director's perspective. Jeremy Manternach (Div. of Music Education and Music Therapy, School of Music, Univ. of Kansas, 1530 Naismith Dr., Lawrence, KS 66045, jmanter@ku.edu)

What do vocal/choral instructors value in a rehearsal space? Along with logistical considerations, these instructors also desire particular room acoustic conditions. How might their input affect the decisions of architects and acousticians? This question is investigated through examination of several school rehearsal spaces. The study will include acoustical measurements from rehearsal spaces, opinions of vocal music instructors, and feedback from vocalists who rehearse in each space.

9:25

3aAA5. Acoustical compromises: Working with a design team for maximum isolation at minimal costs in high school music facilities. Pamela J. Harght (4006 Speedway, Austin, TX 78751, pam@baiaustin.com)

Texas Public High Schools place a large emphasis on their music and drama programs, often providing facilities for their students that are far superior than many 4-year college and universities. These high school facilities include several rehearsal rooms for band, choir, orchestra, percussion, ensembles, mariachi and individual practice. In addition, they also often have a drama classroom, a black box theater, and a Performing Arts Center. When working with the design team on these projects, there are often many obstacles to work

around in order to provide the best scenario possible for the end user. These "obstacles" often include (but are not limited to) school district codes, budgets, the "convenience" factor, city codes, and space planning. This paper will showcase many high school music facilities through the south and southeast and the compromises and obstacles that were overcome with the end result goal as satisfying the music department.

9:45

3aAA6. Helping clients realize superior room acoustics while achieving other facility goals. Rob Farion (Farion Architectural Acoust., 3749 43A Ave., Red Deer, AB T4N 3G2, Canada, rob@farion.ca)

Addressing the needs of a space often involves considering many competing objectives. When room acoustic performance is one of the priorities, there is potential to address other priorities in conjunction with acoustic considerations. Integrating interior design goals with acoustic requirements often allows more resources to ultimately be available for acoustic considerations. This presentation outlines two educational facility projects where clients were assisted in meeting their goals by incorporating interior design objectives and room furnishings with acoustic performance outcomes.

10:05—10:20 Break

10:20

3aAA7. The evolution of the gymatorium and cafetorium in primary schools. Felicia Doggett (Metropolitan Acoust., LLC, 40 W. Evergreen Ave., Ste. 108, Philadelphia, PA 19118, felicia@metropolitanacoustics.com)

Many primary schools do not have formal auditoriums, but the school still needs a space for performances and assemblies. The existence of "gymatoriums" and "cafetoriums" in these schools is essential to the student body and staff. Far from being just a gymnasium or cafeteria with a stage at one end, these rooms are fully functioning performance spaces with appropriate acoustics and audio systems. What many in the school may deem most important, however, is that they still function very well for their everyday uses of gymnasiums or cafeterias. This presentation explores several gymatoriums and cafetoriums and their acoustical and audio design elements.

10:40

3aAA8. Acoustical design of special purpose rooms in schools: Music rooms. Gary W. Siebein (Univ. of Florida School of Architecture, P.O. Box 115702, Gainesville, FL 32611), Hyun Paek, Chris, P. Jones, and Reece Skelton (Siebein Assoc., Inc., Gainesville, FL 32607)

Design strategies for music education rooms in elementary, middle, and high schools based on an impulse response-based theory are presented through a series of case studies of existing, renovated, and designed rooms. The strategies include providing sound reflections to allow the instructor to hear individual groups of students as well as to allow the students to hear each other, controlling room volume and absorption to allow early sound reflections for clear hearing, and simultaneously provide diffuse, low-level running reverberance to enhance musical qualities, providing for clear verbal communication between teacher and students and limiting background and intruding noise levels. The impulse response-based measures are reduced to a series of architectural systems that can be implemented using alternate construction systems to meet budget requirements. Standard design concept sketches communicate the intent of the acoustical design to Architects, School Board members, Design Builders and music faculty.

Session 3aABa

Animal Bioacoustics: Echolocation and Habitat Sonar

Peter Marvit, Cochair
Univ. of Maryland, Dept. of Psychology, College Park, MD 20742

Susan E. Parks, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0300

Contributed Papers

8:00

3aABa1. Modeling of precise onset spike timing for echolocation in the big brown bat, Eptesicus fuscus. Jason E. Gaudette (Ctr. for Biomedical Eng., Brown Univ., Providence, RI 02912, jason_gaudette@brown.edu) and James A. Simmons (Brown Univ., Providence, RI 02912, james_simmons @brown.edu)

The neuronal circuits enabling highly accurate sonar performance in bats are not well understood. In this research, a computational neuroscience model was developed based on Meddis' auditory peripheral model and a biologically plausible network of integrate-and-fire neurons. The stimuli presented to the model include acoustic simulations of target echoes using wideband FM echolocation calls recorded from Eptesicus fuscus. Synaptic connectivity of the neural network was varied to explore its effect on the precision of onset spike timing. Results show that precisely timed spike codes can relay sufficient information of echo time-frequency representations to account for performance observed in laboratory experiments. This research represents a portion of our overarching goal to reverse engineer the bat's critical auditory functions into a refined sonar signal processing algorithm. A necessary first component of this process is to understand the functionality of relevant circuitry in the auditory midbrain. Such biomimetic processing schemes may be useful in active sonar applications suffering from a high degree of environmental clutter, where man-made systems typically demonstrate poor performance.

8:15

3aABa2. Similarities in the echolocation beam pattern of vespertilionid bats. Lasse Jakobsen and Annemarie Surlykke (Inst. of Biology, Univ. of Southern Denmark, Campusvej 55, Odense C, Denmark)

Echolocating bats can navigate and fourage by sound, emitting short high-frequency sound pulses and listening to echoes reflected from obstacles and prey. Small insectivorous bats typically emit higher frequencies than larger bats. This correlation has been explained as an acoustic constraint by prey size since small bats feed on small prey and small objects only reflect echoes efficiently at short wavelengths, i.e., high frequencies. However, very small bats use higher frequencies than what is needed for effective reflection from their prey. Thus, we propose an alternative or supplementary hypothesis: Small bats emit high frequencies to obtain highly directional biosonar sound beams. Directionality increases with the size of the sound emitter relative to the wavelength. Hence small bats can counteract the decrease in directionality caused by their small size by increasing their emitted frequency. Echolocation calls were recorded from five species of echolocating bats of different sizes flying in the laboratory using a multimicrophone array and their beam pattern was calculated. The results show high similarity in beam patterns across species, indicating that directionality is indeed a constraint on echolocation frequency. [Funded by the Oticon Foundation.]

8:30

3aABa3. A comparison of biosonar beampatterns in bats. Rolf Müller (Dept. of Mech. Eng, Virginia Tech & Inst. for Adv. Learning and Res., 150 Slayton Ave., Danville, VA 24540, rolf.mueller@vt.edu) and Zhen Yan (Shandong Univ., 250100 Jinan, China)

The beampattern associated with a diffracting baffle around a sound receiver can be predicted efficiently using numerical methods. Here, such methods have been used to derive estimates for a large number of reception beampatterns from high-resolution digital models of the shapes of pinnae (outer ears) of different bat species. For each ear, beampatterns were obtained with a high-directional resolution (1 deg) and for a set of frequencies that sample the frequency band used by the respective species adequately. This allows the characterization of frequency-dependent features of the beampattern such as the number of lobes present, as well as their size, shape, orientation, and position. The resulting data set allows comparisons between different ecological groups, such as the continuous frequency (cf-fm) and frequency modulated bats (fm-bats) as well as an assessment of variability within such species groups. Comparisons across a larger number of species have facilitated the identification of outliers, i.e., species with unique beampatterns. Such unique beampattern features may be related to unique uses of the animals biosonar system or unique constraints under which these tasks have to be accomplished. Tentative links to unique features of the pinna geometry have also been established.

8:45

3aABa4. Ear deformations in the biosonar system of horseshoe bats. Li Gao, Weikai He (School of Phys., Shandong Univ., Hongjia Lou 5, 250100 Jinan, China, gl1231@mail.sdu.edu.cn), Sreenath Balakrishnan, and Rolf Müller (Virginia Tech. & Inst. for Adv. Learning and Res., Danville, VA 24540)

Many bat species exhibit conspicuous ear movements as part of their biosonar behaviors. The resulting transformations of the pinnae can contain rigid rotations as well as nonrigid deformations. Characteristic ear deformations are frequently observed in horseshoe bats (*Rhinolophidae*). In these animals, the pinna changes regularly between an upright and a bent configuration. In the bent configuration, the tip of the pinna tilts forward, down, and sideways, changing the curvature of large portions of the ear surface. Entire deformation cycles can be completed in as little as 200 ms. The time-variant geometry of the pinna deformations in the greater horseshoe bat (*Rhinolophus ferrumequinum*) was captured with a stereo pair of high-speed video cameras. Landmark points were marked on the pinna with dyes that were equally visible in the video frames as in x-ray micro-CT. The CT model was matched to the locations of the landmarks in the deforming ears inferred from stereo reprojection. Numerical predictions were obtained for the acoustic behavior of ear shapes at different phases during the deformation cycle. It was found that the pinna deformations seen in the animals can produce large changes in the directivity patterns. [Work supported by NSFC (Project No. 10774092) and CSC.]

9:00

3aABa5. A digital model for the deformation of bat ears. Sreenath Balakrishnan (Dept. of Mech. Eng., Virginia Tech. & Inst. for Adv. Learning and Res., 150 Slayton Ave., Danville, VA 24540, bsreenat@vt.edu), Li Gao, Weikai He (Shandong Univ., 250100 Jinan, China), and Rolf Müller (Virginia Tech & Inst. for Adv. Learning and Res., Danville, VA 24540)

In bats, the directivity patterns of the biosonar system are shaped by the surface geometry of the pinnae. Since many bat species are capable of large ear deformations, these beampatterns can be time-variant. To investigate this time-variance using numerical methods, a digital model that is capable of representing the pinna geometry during the entire deformation cycle has been developed. Due to large deformations and occlusions, some of the surfaces relevant to sound diffraction may not be visible and hence the geometry of the entire pinna has to be computed from limited data. This has been achieved by combining a complete digital model of the pinna in one position with time-variant sparse sets of three dimensional landmark data. The landmark positions were estimated using stereo vision methods. A finite element model based on elasticity was constructed from CT scans of the pinna *post mortem*. This elastic model was deformed to provide a good fit to the positions of the landmarks and retain values of smoothness and surface energy comparable to life. This model was able to handle ratios of data to degrees of freedom around 1:5000 and still effect life-like deformations with an acceptable goodness of fit.

9:15

3aABa6. Reanalyzing auditory sensitivity: The functional audiogram as modeled by the bat detecting moth ear. Matthew E. Jackson, Navdeep Asi, and James H. Fullard (Dept. of Biology, Univ. of Toronto at Mississauga, 3359 Mississauga Rd., N. Mississauga, ON L5L 1C6, Canada, matthereric.jackson@utoronto.ca)

Auditory sensitivity has often been measured by identifying neural threshold in real time (online) which can introduce bias in the audiograms that are produced. This was tested by recording auditory nerve activity of the Notodontid moth *Nadata gibbosa* elicited by bat-like ultrasound and analyzing the response offline. The offline audiogram was compared to a published online audiogram showing that the bias introduced can result in a difference in both the best frequency of this moth ear and the audiogram shape. The offline neural threshold audiogram was then compared to audiograms produced using behavioral threshold definitions based on (1) spike period and (2) the latency to first spike, showing that audiograms produced using these theoretical behavioral definitions to have a similar shape. Finally, predictions on the distance at which Notodontid moths may evade bats using negative phonotaxis or the acoustic startle response are made, using the number of spikes elicited as a proxy for distance.

9:30

3aABa7. Echolocation in a fresh water lake. Elizabeth vonMuggenthaler (Fauna Commun. Res. Inst., Hillsborough, NC, 27278 1@animalvoice.com), Joseph Gregory (NC), and Scott H. Mardis (Fauna Commun. Res. Inst., Vermont, Canada)

Lake Champlain is host to myriad interesting geological peculiarities, including the oldest Middle Ordovician reef bed in the world containing coral. Autogenic succession suggests animals adapt to modifications in climate. Eleven thousand year old Beluga whale skeletons have been found in Vermont and Beluga whales currently survive in the St. Lawrence Seaway. In 2003, 2005 and 2009 sites on Lake Champlain were explored using computers with National Instruments Polynesia real-time sound analysis, NI PCMCIA 6062-E cards, DAT recorders, GPS, amplifiers, three vector sensors, two hydrophones, and a Nagra IV-S5. ECHO Aquarium in Vermont facilitated recording of known lake inhabitants and non-biological signals studied formed the basis for the control. Neither recordings nor literature indicate that the known native creatures echolocate. Combining wavelet applications, aiding in reduction in ambient noise in this opposing environment along with conventional analysis, the experiments have been able to conduct far reaching, low-noise sound measurements and were capable of detecting signals the nature of which suggests the presence of some interesting and unexpected phenomena within the ranges and inherent structure of Beluga whale, killer whale, and dolphin echolocation. To protect Lake Champlain, further investigations into this acoustic anomaly is encouraged. [Work supported by the Radius Foundation.]

9:45

3aABa8. Identifying manatee location using dual-frequency sonar DIDSON. Christopher Niezrecki (Dept. of Mech. Eng., UMass Lowell, 1 Univ. Ave., Lowell, MA 01854)

Over the last several years there has been much interest in detecting the West Indian manatee (*Trichechus manatus latirostris*) in turbid waters. Even in clear waters, manatees can be difficult to detect at appreciable distances by looking from above at the surface of the water. The standard DIDSON sonar is a multibeam sonar that provides dynamic images and operates at two frequencies (1.1 and 1.8 MHz). It is capable of providing high-resolution images of objects underwater that are typically scanned from the side or from above. Within this work, the DIDSON sonar is used to detect manatees at Homosassa Springs, FL. The manatees are ensonified from the side and appear to be viewed from the top. The manatees are readily apparent at a range up to 23 m for the measurements acquired in this preliminary sample. The presentation will present some of the images of manatees using a handheld diver DIDSON unit and describes some of the challenges and implications in using the device to detect manatees.

10:00

3aABa9. Using active acoustics to assess habitat restoration in a freshwater lake. Laura E. Madden (Dept. of Wildlife and Fisheries Sci. and Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804-0030, lem230@psu.edu) and Jennifer L. Miksis-Olds (The Penn State Univ., State College, PA 16804-0030)

Artificial structures increase bottom complexity in man-made reservoirs and aim to increase fishery production by providing improved refuge, forage, and reproduction habitat. Electrofishing is typically employed to assess the effectiveness of these structures but is limited to shallow water sites. Active acoustic technology is a non-invasive sampling method that is more flexible than electrofishing, as it is operational in both deep and shallow water sites. Active acoustics also provides a continuous data series at one location to obtain high-temporal resolution information. An acoustic water column profiler was deployed for 1 week at each of two sites: (1) a control site with no introduced structure and (2) a treatment site with artificial refuge habitat. Difference in fish abundance, vertical distribution, and diurnal behavior between the two sites was assessed from the volume backscatter time series. Variation in fish activity between the two sites was compared. This work demonstrates the utility of active acoustics in assessing the effectiveness of freshwater habitat alteration beyond the scope of conventional techniques. A more comprehensive evaluation of habitat restoration is crucial in guiding the development of future conservation efforts.

Session 3aABb

Animal Bioacoustics: Animal Hearing and Vocalization

Susan E. Parks, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Peter Marvit, Cochair
Univ. of Maryland, Dept. of Psychology, College Park, MD 20742

Contributed Papers

10:30

3aABb1. Aerial audiograms of Steller and California sea lions measured using auditory steady-state response methods. Jason Mulsow (U.S. Navy Marine Mammal Program, SSC Pacific, Code 71510, 53560 Hull St., San Diego, CA 92152, jmulsow@hotmail.com), Colleen Reichmuth (Univ. of California, Santa Cruz, Santa Cruz, CA 95060), Frances Gulland (The Marine Mammal Ctr., Sausalito, CA 94965), David A. S. Rosen (Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada), and James J. Finneran (U.S. Navy Marine Mammal Program, SSC Pacific, San Diego, CA 92152)

Detection of aerial vocal signals by conspecifics is important in the reproductive behavior of the otariid pinnipeds. However, aerial hearing sensitivity measurements have only been obtained for a few otariid individuals that were trained to participate in behavioral experiments. In order to expand upon this small data set, auditory steady-state response (ASSR) methods were used to examine the aerial hearing sensitivity of Steller and California sea lions. Although ASSR thresholds were elevated relative to behavioral thresholds reported for otariids, the ASSR audiograms of the majority of individuals were similar to each other and to behavioral audiograms in terms of relative sensitivity. A marked reduction in sensitivity with increasing frequency regularly occurred between 16 and 32 kHz, indicating a consistent high-frequency cutoff. The reliability of the ASSR audiograms for both species suggests that behavioral aerial audiograms that exist for a few Steller and California sea lion individuals can be appropriately extrapolated to larger populations. The similarity of the ASSR audiograms among the Steller and California sea lions supports the notion that the otariid pinnipeds form a functional hearing group, with similar aerial hearing in terms of sensitivity and frequency range of hearing. [Work supported by ONR and NOAA Ocean Acoustics Program.]

10:45

3aABb2. Marine mammal auditory evoked potential measurements using swept amplitude stimuli. James J. Finneran (US Navy Marine Mammal Program, SSC Pacific, 53560 Hull St., San Diego, CA 92152, james.finneran@navy.mil)

Auditory evoked potentials are routinely used to characterize hearing in marine mammals for whom behavioral testing is not practical. For frequency-specific audiometry, the most popular evoked potential method has been the measurement of the auditory steady state response (ASSR). These tests normally entail measuring the ASSR to a sequence of sinusoidally amplitude modulated tones so that the ASSR input-output function can be defined and the auditory threshold estimated. In this study, an alternative method was employed, where thresholds were estimated in response to a single amplitude modulated stimulus whose sound pressure level varied linearly with time. The tone sound pressure was therefore swept over a range of levels believed to bracket the threshold. The input-output function was obtained by analyzing the resulting grand average evoked potential using a short-time Fourier transform. The swept amplitude technique was performed with bottlenose dolphins and California sea lions, with the resulting thresholds similar to those obtained with constant amplitude tones. The tradeoffs between the swept amplitude technique and the traditional, constant amplitude approach will be discussed. [Work supported by ONR.]

11:00

3aABb3. Relative salience of acoustic features in zebra finch song. Beth A. Vernaleo, Robert J. Dooling (Dept. of Psych., Neurosci. and Cognit. Sci. Program, Univ. Maryland, College Park, MD 20740, bgoldman@umd.edu), and Marjorie R. Leek (Natl. Ctr. for Rehab. Aud. Res., Portland, VA and Medical Ctr., Portland, OR 97239)

Zebra finch song consists of complex acoustic elements repeated over several hundred milliseconds with extreme vocal-motor precision. Much less is known about how song is perceived though previous work has shown that zebra finches easily discriminate between normal and time-reversed versions of song syllables. Here, we used operant conditioning, psychophysical methods, and various synthetic song models to assess the bird's sensitivity to local versus global temporal changes in song and to determine which cues are most salient in these complex syllables. Birds could discriminate syllable reversals in songs made up of the song envelope filled with noise, suggesting that syllable envelope cues can provide the basis for discrimination. However, birds could also discriminate syllable reversals in songs made up entirely of Schroeder harmonic complexes, stimuli that provide only phase (reversals of fine structure) cues without syllable envelope cues. Birds performed better on longer syllables, suggesting a window of temporal integration for fine structure discrimination. In contrast, humans could perceive few of the syllable reversals within these song modifications. This fine-grained perception in birds suggests that the machinery used for song perception is as precise as the machinery used for song production. [Work supported by NIH.]

11:15

3aABb4. Distance perception by non-territorial zebra finches (Taeniopygia guttata) and budgerigars (Melopsittacus undulatus). Kelly E. Radziwon and Micheal L. Dent (Dept. of Psych., Univ. of Buffalo, SUNY, 206 Park Hall, NY 14260, radziwon@buffalo.edu)

Birds use long-range acoustic signals to defend territories, attract mates, and locate conspecifics. However, long-range acoustic signals progressively degrade during their transmission from the signaler to the receiver. This degradation makes perceiving these signals more difficult, but receivers can use such information to estimate the signaler's distance. The perception of auditory distance cues (overall amplitude, frequency-dependent attenuation, and reverberation) has typically been studied in the field with territorial birds. The present study examined auditory distance perception in a non-territorial songbird, the zebra finch, and in a non-songbird, the budgerigar in a controlled laboratory setting. Three zebra finches and three budgerigars were trained to identify 1-m (undegraded) and 75-m (degraded) recordings of three budgerigar contact calls, one male zebra finch song, and one female zebra finch call. Test stimuli were then introduced on 20% of the total number of trials. These stimuli were created manually by manipulating the naturally recorded 1-m calls. By editing these calls, we could manipulate each distance cue separately to determine which cue was the most salient for the birds. Our results suggest that amplitude was the most important cue for these birds, similar to humans and other animals.

11:30

3aABb5. Direct measurements of subjective loudness in a bottlenose dolphin. Carolyn E. Schlundt (ITT Corp., 2376 Rosecrans St., San Diego, CA 92106, carolyn.melka@itt.com) and James J. Finneran (US Navy Marine Mammal Program, Space and Naval Warfare Systems Ctr. Pacific, San Diego, CA 92152)

Equal-loudness contours were measured in a bottlenose dolphin (*it Tursiops truncatus*) trained to perform a loudness comparison task. The subject was presented two sequential tones and whistled if the first tone was louder, and produced a burst pulse or "buzzed" if the second tone was louder. Approximately 70% of trials were "known" comparisons [e.g., tones of same frequency but different sound pressure levels (SPLs)] and allowed performance to be tracked within sessions. The remaining comparisons were probe trials, consisting of a 10-kHz standard tone with a fixed SPL (either 90, 105, or 115 dB *re* 1 μPa) and a comparison tone, whose frequency was fixed but whose SPL varied. Presentation order of probe trials was balanced. Eleven comparison frequencies ranged from 3.5 to 113.1 kHz. Logistic regression was used to derive curves relating the probability of the comparison tone being perceived louder. The 50% point represented the SPL at which the comparison and standard tones were equally loud. The data represent the first direct measurement of equal-loudness curves in any animal and show the relationship between the frequency and subjective loudness. Loudness contours may be more appropriate for assessing behavioral effects of sound, assuming behavioral reactions are more strongly related to *loudness* than *SPL*.

11:45

3aABb6. Vocalization in a neonatal Amur tiger cub. Edward J. Walsh, Adam B. Smith (Developmental Auditory Physio. Lab., Boys Town Natl. Res. Hospital, 555 North 30th St., Omaha, NE 68131, smithab1@boystown.org), Douglas L. Armstrong (Omaha's Henry Doorly Zoo, Omaha, NE 68107), and JoAnn McGee (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

Although the acoustical features of a small number of calls representing the vocal repertoire of adult tigers have been studied and reported, calls produced by developing tigers representing any subspecies have not been described. In that light, the acoustical features of calls produced by an Amur cub during the first two postnatal weeks of life will be considered. The vocal repertoire of the cub observed during recording sessions was limited, consisting primarily of neonatal cries composed of three distinct types: an intense, harmonically rich phonation, an aperiodic cry (dysphonation), and a compound cry exhibiting both dysphonation and harmonically rich phonation segments. Overall acoustic power of the harmonically rich phonation was uniformly distributed across a frequency band ranging from approximately 0.6–2.0 kHz and that was centered on roughly 1.2 kHz on the second postnatal day. While the same basic pattern was observed on postnatal day 13, inter-harmonic differences were smaller, and between days 2 and 13 the mean fundamental frequency dropped from ~260 to ~200 Hz; typically, f0 varied during each utterance at each age studied. These findings will be considered in relation to the spectrographic character of calls produced by adult Amur tigers. [Funding was provided by NSF Grant 0823417.]

WEDNESDAY MORNING, 21 APRIL 2010 KENT A/B, 7:55 TO 11:30 A.M.

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Numerical Modeling for Medical Ultrasound II

Robert J. McGough, Cochair
Michigan State Univ., Electrical and Computer Engineering, 2120 Engineering Bldg., East Lansing, MI 48824

Vera A. Khokhlova, Cochair
Moscow State Univ., Acoustics Dept., Leninskie Gory, Moscow, 119992, Russia

Chair's Introduction—7:55

Invited Papers

8:00

3aBB1. Numerical modeling techniques common to medical ultrasound and detection of undersea objects. David S. Burnett (Naval Surface Warfare Ctr., Panama City Div., Code T11, 110 Vernon Ave., Panama City, FL, david.s.burnett@navy.mil)

Acoustic scattering from gallstones, diseased tissues, and objects on the ocean floor and submarines are all governed by the same underlying partial differential equations describing wave propagation in fluids and viscoelastic solids. These objects differ in size by many orders of magnitude. However, they are all the same "acoustic size" when each is insonified at a frequency such that the ratio of object-diameter-to-wavelength is the same for all. Objects of the same acoustic size have two important characteristics: (1) they exhibit qualitatively similar physical-acoustic phenomenology and (2) numerical modeling consumes similar computational resources. Given these observations, many of the numerical modeling techniques developed by the U.S. Navy for acoustic scattering from undersea objects may be applicable to biomedical ultrasound applications. This paper explores these possibilities. NSWC PCD has developed a computer simulation system for modeling the acoustic color (target strength versus frequency and aspect angle) of realistic objects. It automatically runs hundreds of thousands of three dimensional FE models, changing the mesh and finite fluid boundaries of the models as they sweep over frequency and yielding uniform accuracy with frequency. This paper will present an overview of the technology and numerical verification and validation of the system. [Work supported by ONR and SERDP.]

8:20

3aBB2. Optimization of a 100 MHz fiber optic hydrophone for medical ultrasound applications using nanometer scale gold coating. Rupa Gopinath, Khushali Manseta, Vasileios Nasis, Afshin S. Daryoush (Dept. ECE, Drexel Univ., 3141 Chestnut St., Philadelphia, PA 19104), Mahmoud El-Sherif (Photonics Labs Inc., Philadelphia, PA 19104), and Peter A. Lewin (Drexel Univ., Philadelphia PA 19104)

This work describes a novel numerical model developed to optimize the design of a broadband, spatial averaging free fiber optic hydrophone probe (FOHP) for characterization of ultrasound fields in the frequency range 1–100 MHz. The design included minimization of the active sensor dimensions, so its cross-section is comparable with the half acoustic wavelength at the highest frequency of interest and improvement of voltage-to-pressure sensitivity of the probe measured in V/Pa. Initial simulations based on the assumption that bulk refractive index of sputtered material could be used indicated that a thin film gold coating (1–35 nm) could indeed enhance the voltage sensitivity of the FOHP by 16–30 dB; however, a follow up analysis revealed that determination of the coating thickness influence on the FOHP performance would require an introduction of the complex (as opposed to bulk) index of refraction of the sputtered film. The input parameters to the model, their selection criteria, and implementation of the coupled acousto-optic interaction of gold layer will be discussed. The measured prototypes produced unprecedented voltage sensitivity between −234 and −245 dB *re* 1 V per uPa or 2 and 560 mV/MPa, respectively. [The authors wish to acknowledge support of the NIH R01EB007117 grant.]

Contributed Papers

8:40

3aBB3. Acoustic streaming for direct delivery of therapeutics into tissue. Raghu Raghavan (1101 East 33rd St., Ste. B305, Baltimore, MD 21218)

Ultrasound has been used in medical drug delivery but largely for altering membrane permeability. We are investigating acoustic streaming for delivery of therapeutics, such as drug molecules or other particulates, into brain tissue, with other applications for the future. Acoustic streaming potentially can enhance fluid and particle flux in interstitial pathways in tissue. The conventional equations that describe streaming in fluids are not directly applicable to porous media such as the brain. In this study, a modified Biot equation is employed to develop a theoretical framework to account for the attenuation mechanisms dominant in acoustic streaming in porous media. The analytical solutions obtained for isotropic, homogeneous media are presented and compared with experiments conducted on gel phantoms, and with results reported for streaming in live monkey brains. The dependence of streaming-induced therapeutic transport on acoustic parameters and tissue properties are investigated. The theoretical framework was used to determine optimal operating parameters of ultrasonic devices for streaming-based drug transport in porous tissue with particular emphasis on two objectives: (1) reduction in backflow from catheters inserted into brain tissue for direct delivery and (2) delivering drugs into the margins of a resection cavity following surgery for brain tumor.

8:55

3aBB4. Phase synchronization and collective instability in oscillating bubble clouds. D. Sinden, E. Stride, and N. Saffari (Dept. Mech. Eng., Univ. College London, Torrington Pl., London WC1E 7JE, United Kingdom, d.sinden@ucl.ac.uk)

The effect of bubble interactions within a bubble cloud under ultrasonic forcing is investigated. Under certain conditions, phase synchronization is found, revealing parameter regimes where the dynamics of the bubble system are primarily determined by the interactions of the bubbles, rather than the applied ultrasonic acoustic field. Due to the Bjerknes interaction, the stability thresholds for critical pressure and size leading to unpredictable oscillations for bubble clouds are lower than the stability thresholds for single bubbles. Interestingly, numerical simulations confirm that in the regimes where bubbles exhibit chaotic oscillations, the dynamics are determined by a complicated geometric object called a low-dimensional strange attractor which is qualitatively the same as that for the chaotic oscillations of a single bubble. Thus an averaged equation may also be low-dimensional. The effects of synchronization on Landau damping are discussed. The effect of delays between the interactions is also investigated. The implications for applications in high-intensity focused ultrasound, harmonic imaging, and targeted drug delivery will be discussed.

9:10

3aBB5. Design and implementation of a cylindrical array for passive mapping of cavitation fields produced by clinical high intensity focused ultrasound transducers. Stuart Faragher, Miklós Gyöngy, Jamie Collin (IBME, Univ. of Oxford, ORCB, Oxford OX3 7DQ, United Kingdom), Mark Hodnett (Quality of Life Div. Natl. Physical Lab., Teddington TW11 0LW, United Kingdom), and Constantin-C. Coussios (Univ. of Oxford, Oxford OX3 7DQ, United Kingdom)

Key to the success of high-intensity focused ultrasound (HIFU) as a clinical tool is the development of standardized quality assessment procedures to assess the safety and efficacy of HIFU transducers. The present work details the development of a cylindrical sensor array to be positioned around the HIFU focus during pre-treatment quality assessment of clinical transducers, which is designed to localize cavitation activity in three dimensions by passive mapping of the broadband emissions arising from inertial cavitation. The propagation of sound from a collection of broadband sources was first modeled to determine the optimum size, number and distribution of array elements for accurate mapping, and characterization of the cavitation dynamics produced during HIFU exposure. The optimal array configuration was then manufactured from PVDF using a novel printed circuit board technique, and theoretical predictions of the spatial resolution that it could achieve were validated experimentally. Because inertial cavitation is a pressure driven phenomenon, the ability to map cavitation activity using this array could provide a novel tool for rapid mapping of pressure fields produced by clinical HIFU transducers, in addition to providing invaluable information about the evolution of cavitation dynamics during HIFU exposure.

9:25

3aBB6. Diffraction pattern recognition method for drastically improving image resolution and diagnostic sensitivity. J.J. Wen (VeroSound, Inc., 128 Powder Mill Rd., Sudbury, MA 01776)

Compared with the two other mainstream medical imaging methods, CT and MRI, the worst weakness of ultrasound imaging is poor resolution—the ability to resolve tiny variations in tissue structure or texture. Due to sound wave diffraction and the way imaging ultrasound signal is retrieved, a point target in object domain does not generate point image in image domain, but an image spot with sophisticate spot pattern and considerable spot size as functions of point target location relative to the transducer—the most direct cause of poor resolution. The mainstream techniques aimed at reducing the image spot, either by shortening the impulse signal or by sharpening the beam focusing, are unavoidably accompanied by the deteriorated sensitivity and depth of penetration. It is a common consent in medical ultrasound community that ultrasound imaging is very close to its theoretical resolution limit. This article presents a different approach we named E-mode imaging that uses a diffraction-theory-based spot pattern recognition technique to account for the effects of image spot pattern in image processing. With the same set of ultrasound data that *B* imaging is based on, E-mode imaging achieved five to ten times better resolution and diagnostic power in computer simulations.

9:40

3aBB7. A generalized formulation for scattering from arbitrary distributions of homogeneous cylinders or spheres. Andrew J. Hesford, Jeffrey P. Astheimer, and Robert C. Waag (Dept. of Elec. and Comput. Eng., Univ. of Rochester, UR Med. Ctr. Box 648, 601 Elmwood Ave., Rochester, NY 14642-8648, hesford@ece.rochester.edu)

A generalized form of spatial harmonic expansion suitable for calculation of scattering by arbitrary distributions of homogeneous cylinders in two dimensions and homogeneous spheres in three dimensions is presented. The method is applicable for general incident fields expressed as angular spectra of plane waves. The technique uses harmonic expansions, physical acoustic boundary conditions, and mode matching to represent scattering by each individual object as a diagonal operator. Interaction between objects is treated with a *T*-matrix approach in which scattering from one object is applied to another using translation operators that shift the origins of harmonic expansions between scatterers. The generalized form permits efficient calculations in both two and three dimensions. These calculations are used to assess the validity of two-dimensional approximations to three-dimensional scattering for volumes containing multiple interacting spheres that are illuminated by elevation-focused beams. Differences in the two-dimensional and three-dimensional results are attributed to out-of-plane scattering and partial-volume effects caused by significant variations in the scatterer cross section over the elevation width of the focused beam.

9:55—10:30 Break

10:30

3aBB8. Comparison of the Khokhlov–Zabolotskaya–Kuznetsov equation and Fourier-continuation for modeling high-intensity focused ultrasound beams. Theresa Cheung, Robin O. Cleveland (Dept. of Mech. Engin., Boston Univ., Boston, MA 02215), Nathan Albin, and Oscar P. Bruno (Caltech, Pasadena, CA 91125)

High-intensity focused ultrasound (HIFU) employs acoustic amplitudes that are high enough that nonlinear propagation effects are important in the evolution of the sound field. A common model for HIFU beams is the Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation which accounts for nonlinearity and absorption of sound beams. The KZK equation models diffraction using the parabolic or paraxial approximation. For many HIFU sources the source aperture is similar to the focal length and the paraxial approximation may not be appropriate. Here we compare results obtained using the "Texas code" a time-domain numerical solution to the KZK equation to a new method of solving nonlinear differential equations: the Fourier-continuation (FC) method. Here the FC method solves the underlying fluid dynamics equations and does not invoke the paraxial approximation. Results were obtained for a 1.1-MHz focused HIFU source. The medium parameters were taken to either be hyperviscous water or to match human liver. For a low focusing gain transducer (focal length 50λ and radius 10λ) the KZK and FC models showed excellent agreement through the focal region. As the source radius was increased, discrepancies started to appear and for a radius of 30λ, the KZK model did not capture diffraction of the HIFU source accurately. [Work supported by NSF 0835795.]

10:45

3aBB9. Characterization of clutter in ultrasonic imaging using a nonlinear full-wave simulation method and *in vivo* human tissue. Jeremy J. Dahl (Dept. of Biomedical Eng., Duke Univ., Durham, NC 27708, jeremy.dahl@duke.edu) and Gianmarco F. Pinton (Laboratoire Ondes et Acoustique, ESPCI, Paris, France)

Clutter is a well known image degradation mechanism in ultrasonic imaging that is patient dependent. Despite widespread agreement on clutter sources, there are few studies on sources and their contribution to overall image clutter. We have performed simulations of clutter using a nonlinear, full-wave method to demonstrate the characteristics and sources of ultrasonic clutter. Representations of human abdominal layers, created from histological stains of human abdominal layers, were placed at the surface of the transducer. Point targets and anechoic cysts were simulated to demonstrate the characteristics of clutter resulting from these near-field layers. We have also performed matching *in vivo* experimental studies in human bladders and livers. The simulations and experiments show consistent results. Clutter resulting from near-field tissue layers is spatially coherent with respect to axial motion and incoherent in the sampled pressure field. The contribution of off-axis clutter resulting from the tails of the point-spread-function is shown to be relatively small ($<5\%$) compared to the contribution of clutter associated with near-field layers (approximately 68%). Thus, the dominating source of clutter is associated with multiple reflections in near-field tissue layers. Clutter due to off-axis scattering dominates only when an interfering target is close to the imaging target. [This work is supported by the NIH Grant R21-EB008481 from the NIBIB. Technical support was provided by the Ultrasound Division at Siemens Medical Solutions USA, Inc.]

11:00

3aBB10. Modeling nonlinear acoustic waves in media with inhomogeneities in the coefficient of nonlinearity. L. Demi (Lab. of Acoust. Imaging and Sound Control, Fac. of Appl. Sci., Delft Univ. of Technol., Lorentzweg 1, 2628 CJ Delft, The Netherlands, l.demi@tudelft.nl), M. D. Verweij (Delft Univ. of Technol., 2628 CD Delft, The Netherlands), and K. W. A. van Dongen (Delft Univ. of Technol., 2628 CJ Delft, The Netherlands)

The refraction and scattering of nonlinear acoustic waves play an important role in the realistic application of medical ultrasound. One cause of these effects is the tissue dependence of the nonlinear medium behavior. A method that is able to model those effects is essential for the design of transducers for novel ultrasound modalities. Starting from the Westervelt equation, nonlinear pressure wave fields can be modeled via a contrast source formulation, as has been done with the INCS method. An extension of this method will be presented that can handle inhomogeneities in the coefficient of nonlinearity. The contrast source formulation results in an integral equation, which is solved iteratively using a Neumann scheme. The convergence of this scheme has been investigated for relevant media (e.g., blood, brain, and liver). Further, as an example, the method has been applied to compute the one-dimensional nonlinear acoustic wave field in an inhomogeneous medium insonified by a 1 MHz Gaussian pulse propagating up to 100 mm. The results show that the method is able to predict the propagation and the scattering effects of nonlinear acoustic waves in media with inhomogeneities in the coefficient of nonlinearity. This motivates a similar extension of the three-dimensional INCS method.

11:15

3aBB11. Three-dimensional calculation of ultrasound propagation through a breast model. Jason C. Tillett (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Medical Ctr. Box 648, Rochester, NY 14642, tillett@ece.rochester.edu), Gheorghe Salahura, Leon A. Metlay, and Robert C. Waag (Univ. of Rochester, Rochester, NY 14627)

A specimen of human breast comprised mostly of fat and connective tissue was imaged with a high-resolution magnetic resonance scanner. The volume of data was segmented to construct an acoustic model by assigning values of sound speed, density, and attenuation to voxels corresponding to fat and connective tissue. Propagation of a wavefront from a point source through the model was calculated using a three-dimensional finite-difference time-domain *k*-space method that included perfectly matched absorbing boundary conditions to eliminate reflections at the edge of the computational domain. Because of the large $557\times451\times289$ grid and computational demands of the three-dimensional FFT, the calculation was performed on a cluster of computers. The wave field was sampled in three orthogonal planes, one of which was offset from the origin to represent a receiving aperture. The aperture waveforms were corrected for geometry associated with the estimated location of the point source and used to calculate arrival-time and energy-level fluctuations. The root-mean-square arrival-time and energy-level fluctuations of the simulated propagation were consistent with measurements for comparable specimens of breast. The results demonstrate the feasibility of using the described combination of methods to provide insight about the way ultrasound beams are aberrated by tissue.

Session 3aEA

Engineering Acoustics and Psychological and Physiological Acoustics: Acoustic Impedance of the Ear

Daniel M. Warren, Cochair
Knowles Electronics, Inc., 1151 Maplewood Dr., Itasca, IL 60143

Susan E. Voss, Cochair
Smith College, Picker Engineering Program, Northampton, MA 01063-0100

Chair's Introduction—8:00

Invited Papers

8:05

3aEA1. Middle-ear input impedance and middle-ear sound transfer. John J. Rosowski, Hideko H. Nakajima, Jeffrey T. Cheng, Mohamad A. Hamadeh, and Michael E. Ravicz (Eaton-Peabody Lab., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston MA 02114)

Measurements of the acoustic input impedance of the human eardrum, or tympanic membrane (TM), have been used as indicators of middle-ear sound transfer for over 70 years. Modern measurements include (i) middle-ear input impedance in single- and multi-frequency tympanometry, (ii) impedance-based estimates of ear-canal reflectance, a description of sound power absorption at the TM, and (iii) laser-Doppler measurements of the sound-induced velocity of specific points on the TM surface. The advantages and disadvantages of these techniques are discussed, and their use in identifying and estimating the effects of different middle-ear pathologies is illustrated. A new technique, opto-electronic holography, which allows measurements of the volume displacement of the visible surface of the TM, is also discussed. Common themes are (i) the impedance of the middle ear varies widely within populations of normal hearing subjects and (ii) input-impedance based estimates of sound transfer through the middle ear are hindered by the presence of flexibility within the ossicular chain. It is concluded that while middle-ear input impedance may adequately describe the acoustic load on hearing-aid receivers, this impedance, by itself, is not always an accurate indicator of pathological alterations in sound transfer through the human middle ear.

8:25

3aEA2. Effects of middle-ear disorders on ear-canal reflectance measures in human cadaver ears. Susan E. Voss (Picker Eng. Program, Smith College, Ford Hall, 100 Green St., Northampton, MA 01063, svoss@smith.edu), Gabrielle R. Merchant, and Nicholas J. Horton (Smith College, Northampton, MA 01063)

The development of acoustic reflectance measurements may lead to noninvasive tests that provide information currently unavailable from standard audiometric testing. Few systematic measurements exist on the effects of various middle-ear pathologies on the ear-canal reflectance, where reflectance is calculated from impedance measurements. This work uses a human cadaver preparation to examine how a variety of middle-ear pathologies affect power reflectance: static middle-ear pressures, middle-ear fluid, fixed stapes, disarticulated incudo-stapedial joint, and tympanic-membrane perforations. The change from normal in power reflectance, induced by the simulated pathology, generally depends on frequency and the extent of the pathology (except for the disarticulation). The largest changes from normal occur as increases in power reflectance near 1000 Hz with both large static pressures and large amounts of middle ear fluid. Disarticulation of the stapes results in reductions in power reflectance at the lower frequencies, while fixation of the stapes results in low-frequency increases in power reflectance. Tympanic membrane perforations generally produce reductions in power reflectance for frequencies below 1000–2000 Hz that are systematically related to the size of the perforation. These preliminary measurements may help assess the utility of power reflectance as a diagnostic tool for middle-ear disorders.

8:45

3aEA3. Modeling diagnosis of the human middle ear. Pierre Parent (Mimosa Acoust., Urbana IL) and Jont Allen (Dept. of ECE, Univ. of IL, Urbana IL, jontalle@illinois.edu)

Tympanic membrane (TM) and ossicular (OSC) models have been studied for many years. Most are implemented in the frequency-domain via Kirchhoff's equations. However, the frequency domain is less intuitive than the time-domain models and does not apply given non-linear phenomena. This research presents a human time-domain TM+OSC implementation, extending our cat model [J. Acoust. Soc. Am. **122**, 918–931 (2007)] as a cascade of slightly missmatched transmission lines. The canal plane-wave volume velocity is uniformly distributed across the canal cross-section, and updated at $\approx$1 MHz (to accurately simulate sound propagation). The time-domain model includes both forward and backward waves, enabling for the first time, the computation of one-way delays, thus explaining why multi-modal propagation observed on the TM are not critical. Results are validated with data from normal and pathological ears. Additionally, comparisons between the acoustic impedance in living and dead (temporal bones) ears shows significant temporal-bone artifacts. Because of ear-dependent impedance missmatch, canal standing waves are observed. It is frequently assumed that the umbo pressure is the best reference pressure. In fact the "forward-pressure" P_+ is a better estimate of the ideal reference, the cochlear pressure [J. Acoust. Soc. Am. **125**, 1605–1611 (2009)].

9:05

3aEA4. An eardrum impedance simulator for use with physical replicas of the human ear canal. Michael R. Stinson and Gilles A. Daigle (Inst. for Microstructural Sci., Natl. Res. Council, Rm. 307A, Bldg. M-50, Ottawa, ON K1A 0R6, Canada)

An eardrum impedance simulator has been incorporated into life-size replicas of the human ear canal. The simulator branches off from the ear canal at the center of the nominal eardrum position. It has been designed so that sound pressure distributions in an ear canal will have the same standing wave ratios as those measured previously in human subjects over the 4–10-kHz range. The key component of the eardrum simulator is a micro-perforated plate, whose holes provide damping through viscous and thermal boundary layer losses. The optimum design specifies 109 holes of 0.06 mm diameter in 0.003-in. shim stock. Measurements of sound pressure distribution in replica ear canals are underway to evaluate the acoustical performance of the eardrum simulator. The sound field is generated by an occluding hearing aid test fixture.

9:25

3aEA5. Forward-pressure level for in-the-ear calibration. Stephen T. Neely and Michael P. Gorga (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, neely@boystown.org)

When acoustic impedance of the ear canal is known, sound pressure measured by a microphone in the ear canal can be decomposed into forward and reflected components. This decomposition provides a means to overcome problems associated with standing waves that may cause in-the-ear sound level calibration to be unreliable at high frequencies. Compared to total pressure measured at the sound source, we have demonstrated that forward pressure provides a more reliable estimate of sound quantity entering the ear, especially at frequencies above 4 kHz. Specifically, we have demonstrated increased reliability in measurements of distortion-product otoacoustic emissions (DPOAEs) when we use forward-pressure level (FPL) calibrations. Also, we have shown that behavioral threshold measurements for stimuli calibrated in FPL are more reliable than conventional calibration methods in the 4–10 kHz range. FPL calibration may provide advantages at high frequencies (1) for assessing hearing status with DPOAEs and (2) for fitting hearing aids using real-ear measurements. Objective methods of assessing hearing loss and fitting hearing aids at high frequencies are needed because the 6–12 kHz range is known to be important to young children for learning speech. [Work supported by Grants DC2251, DC8312, and DC4662 from the NIH/NIDCD.]

9:45

3aEA6. *In-situ* calibrated sound signals and hearing sensitivity. Robert Withnell (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, rwithnel@indiana.edu), Pat Jeng (Mimosa Acoust., Champaign, IL 61820), William Shofner (Indiana Univ., Bloomington, IN 47405), and Jont Allen (Univ. of Illinois, Urbana, IL 61801)

The acoustic input impedance of the ear provides data on the acoustico-mechanical function of the ear and can be used diagnostically to assess outer and middle ear function. Sound signals delivered to the ear (canal) can be quantified *in-situ* if the complex reflectance of the ear has been determined. Hearing sensitivity can be expressed in terms of the forward-going sound pressure wave, or after correcting for acoustic delay, the fraction of the forward-going sound pressure wave transmitted to the middle ear. The latter is an estimate of the signal the cochlea receives. Data will be presented examining the virtues of expressing hearing sensitivity in terms of an *in-situ* calibrated sound signal.

10:05—10:25 Break

10:25

3aEA7. Simple models for hearing-aid vents and open fittings. James M. Kates (GN ReSound, Dept. Speech Lang Hear Sci., Univ of Colorado, 409 UCB, Boulder, CO 80309)

The effects of a vent or open fitting are determined by the vent and ear acoustics. The dominant effects are a high-pass filter applied to the hearing-aid frequency response and a low-frequency direct signal path through the vent into the ear canal. Because these effects occur predominantly at low frequencies, simple lumped-parameter acoustics can be used to model them. The vent is modeled as an acoustic mass, and the ear canal and middle ear impedances as equivalent acoustic volumes. The results of using these simple approximations will be compared to calculations using acoustic tube models for the vent and ear canal and the Zwislocki middle-ear impedance model.

10:45

3aEA8. A comparison of two methods for estimating sound pressure level at the tympanic membrane at high frequencies for hearing aids. Tao Zhang, Karrie Recker (Starkey Labs., Inc., 6600 Washington Ave. South, Eden Prairie, MN 55344-3405, tao_zhang@starkey.com), Janice LoPresti (Knowles Electrons LLC, Itasca, IL 60143), Matt Kleffner (Starkey Labs., Inc., Eden Prairie, MN 55344-3405), and William Ryan (Knowles Electrons LLC, Itasca, IL 60143)

For hearing aid fittings, it is important to know the delivered sound pressure level (SPL) at the tympanic membrane (TM) over a wide frequency range. Because it is infeasible to measure the SPL at the TM clinically, an indirect estimation based on measurements away from the TM is often used in practice. Furthermore, it is highly desirable that such a method does not require a clinician to use

extra accessories to perform extra calibration in a clinic. Two such methods were proposed in the past. One used a set of empirically derived correction factors to estimate the SPL at the TM (Recker *et al.* (2009), the other used a lumped-element model to estimate the SPL at the TM [LoPresti *et al.* (2009)]. In this study, both methods were evaluated using actual hearing aid prototypes. An ITC hearing aid was built for each of ten participants. For each participant, the SPL in the ear canal was measured using a hearing aid microphone. The SPL at the TM was estimated using each method separately. The results were compared with the measured SPL at the TM. The relative strength and weakness of each method were discussed with the implications for clinical applications.

Contributed Papers

11:05

3aEA9. Assess eardrum characteristics using standing wave technique. Weili Lin and Karrie Recker (Starkey Labs., Inc., 6600 Washington Ave. S., Eden Prairie, MN 55344, weili_lin@starkey.com)

It has been shown that eardrum reflection coefficients can be derived accurately up to 8 kHz based on standing wave patterns in the human ear canal, but not as reliably in higher frequencies [Lawton *et al.*, J. Acoust. Soc. Am. **79**, 1003 (1986)]. The primary purpose of this study is to investigate various factors affecting the accuracy of this type of measurements and extract important characteristics of ear canal and eardrum. In addition to reflection coefficients, eardrum impedance and the effect of loading on the eardrum are also examined. A model ear canal is constructed using a tube with variable loading conditions at one end. Measurements are taken along the center axis and contour of the tube when a stimulus is presented with frequency components up to 16 kHz, under both occluded and open situations. Results obtained from standing wave properties in the model ear canal are then compared to data acquired in real ear and in an ear simulator.

11:20

3aEA10. Capturing low-frequency cochlear impedance in time domain models: The role of viscosity. Michael J. Rapson and Jonathan C. Tapson (Dept. of Elec. Eng., Univ. of Cape Town, Private Bag, Rondebosch, 7701, South Africa, michael.rapson@uct.ac.za)

Viscous effects in the cochlear fluid have been shown to be negligible for frequencies greater than 100 Hz [M. Viergever, Ph.D. thesis]. However, they become important at lower frequencies; hence viscous damping is sometimes included in helicotrema models [S. T. Neely and D. O. Kim, J. Acoust. Soc. Am. **79**, 1472–1480 (1986)]. Modeling viscosity in time-domain computational cochlear models for wide-band input signals is problematic. Since the boundary layer thickness is inversely proportional to frequency, Navier–Stokes formulations require fine spatial resolutions, and fluid potential formulations are inherently inviscid models. Lumping all viscous effects into the helicotrema is not physiologically accurate as "the helicotrema does not appear to be a dominant constriction" [Lynch *et al.*, J. Acoust. Soc. Am. **72**, 108–130 (1982)]. An alternative is proposed whereby a fluid potential formulation is augmented with frequency dependent viscosity corrections that are based on prior analysis of the cochlear geometry and become negligible at higher frequencies. For simplified fluid boundary conditions, the predictions of this fluid model are compared to Navier–Stokes, potential, and lumped viscosity models for frequencies in the range 10 Hz–10 kHz. The simulations are also compared to physiological data for the input impedance of the cochlea.

11:35—11:55 Panel Discussion

WEDNESDAY MORNING, 21 APRIL 2010 KOZMOS, 8:25 A.M. TO 12:00 NOON

Session 3aMU

Musical Acoustics, Signal Processing in Acoustics, and Engineering Acoustics: Measurement and Modeling of the Acoustic Properties of the Banjo

Grace A. Clark, Chair
Lawrence Livermore National Lab., 7000 East Ave., Livermore, CA 94550

Chair's Introduction—8:25

Invited Papers

8:30

3aMU1. What is known about banjo science and what is not yet known. Thomas D. Rossing (Stanford Univ., Stanford, CA 94305, rossing@ccrma.stanford.edu)

Banjos generally come in three differrent types: five-string banjo with a resonator, five-string banjo without a resonator, and four-string banjo with a resonator. Only recently have banjos been the object of much scientific study, and most of these studies have been directed at the American five-string banjo. These investigations have considered the modes of vibration of the principal parts, how they couple together, and how they radiate sound. In general, peaks in the driving point mobility correlate well with modes of vibration in the banjo head and with the radiated sound. Systematic modifications of some of banjo parts have been made by various investigators, and the effects of these modifications are considered.

8:55

3aMU2. A world beyond the banjo. Paul A. Wheeler (Dept. of ECE, Utah State Univ., 4120 Old Main Hill, Logan, UT 84322-4120, paul.wheeler@usu.edu)

The American banjo, a long-necked plucked lute developed by slaves from Africa and an important instrument in Bluegrass music, derives its characteristic timbre from the membrane belly of the instrument. Even though the banjo is the best known membrane-style lute in music of the western hemisphere, there are many musical traditions across the eastern hemisphere which have used their own version of the banjo long before the development of the American banjo. This paper will introduce some of these instruments and the musical cultures using them. These include the ngoni of western Africa (possibly the source of the American banjo), the cumbus and tar from the Middle East, the rawap of central Asia, the sarod of India, and the shamisen and sanxian of Japan and China. As diverse as these cultures may be, the characteristic timbre of the banjo is an important element in their music.

9:20

3aMU3. Time-resolved studies of banjo head motion. Thomas R. Moore and Laurie A. Stephey (Dept. of Phys., Rollins College, Winter Park, FL 32789)

Almost all of the radiated sound from a banjo can be attributed to the motion of the membranic head. This motion depends critically on the physical properties of the membrane as well as which string is plucked and the direction of the initial displacement. We present time-resolved studies of the motion of the membrane of an American five-string banjo after plucking and discuss the interaction between the string and the vibrating membrane.

9:45

3aMU4. The banjo: Developing a plausible model from first principles. Joe Dickey (3960 Birdsville Rd., Davidsonville, MD 21035, dickey@jhu.edu)

The banjo is one of the few musical instruments which is amenable to analytical modeling. It is fundamentally two connected canonical wave bearing systems, and as such, it can be solved. Furthermore, the requisite computer code runs quickly, making it practical to examine the instrument's response as functions of a variety of parameters. The parameters are the things banjo players tweek in their attempt to get the sound they want; things like head tension, bridge mass, and string gauge. This tutorial will start with first principles in both the physics and the mathematics and end with movies of the calculated head response and radiated sound as functions of time and frequency. Numerical models can never be exact. Real systems, and particularly musical instruments, are just too complex for this, but models can provide useful guidance by illuminating what might happen as parameters are changed.

10:15—10:30 Break

10:30

3aMU5. The banjo, string theory for the rest of us. Joe Dickey (3960 Birdsville Rd., Davidsonville, MD 21035, dickey@jhu.edu)

An analytical model of a five string banjo is used to calculate the dynamic response of a plucked banjo and the subsequent radiation of sound. This is the easy part: the difficult part is to identify and quantify the subtle differences between bad and good sound, whether that sound is measured or calculated. As a step toward this, the author has defined figures of merit (FOMs) to quantify radiated sound characteristics related to loudness, brightness, and decay, and to examine these FOMs as functions of the tension and propagation loss factor of the membranic head, the mass of the bridge, and the location of the pluck on the string. The calculated response does exhibit dual decay constants for radiated sound and a diminution of the sixth overtone. Qualitative agreement between calculated and measured mode patterns will be presented.

10:55

3aMU6. Membrane modes and air resonances of the banjo using physical modeling and microphone array measurements. Florian Pfeifle and Rolf Bader (Musikwissenschaftliches Institut, Universitaet Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, Germany, florian.pfeifle@haw-hamburg.de)

A fullscale physical model of a banjo is implemented using a finite-difference formulation. The model consists of strings, bridge, resonance membrane, and the air enclosed in the geometry. The resulting time series of plucked strings in the model radiated from the membrane is integrated with respect to a room position in front of the banjo membrane and analyzed. Furthermore, the modes of the membrane are displayed as two-dimensional mode shapes for the plucked string as well as for the free vibrating membrane. For verification and refinements of the model, measurements on a banjo were performed using a 11×11 microphone array. The recorded time series were back-propagated to the surface of the membrane showing its vibrational patterns. These measured modes are compared to the modeed mode shapes in the low- to mid-frequency range. Furthermore, the air modes of the open sound holes are investigated with respect to the their phase relations to the membrane. Radiation strength of these frequencies are strongly influenced by the opposite phase relations of these patterns. Also open-back and closed-back banjos are compared. The radiation behavior of these lower modes in the far field is then used to be implemented into the physical model of the banjo.

11:20—12:00
A short concert will follow the session in which techniques of Bluegrass 5-string banjos are demonstrated.

Session 3aNCa

NOISE-CON: Plenary

Courtney B. Burroughs, Chair
Noise Control Engineering, Inc., 1241 Smithfield St., State College, PA 16801

Chair's Introduction—8:00

Invited Paper

8:05

3aNCa1. Effects of building mechanical system noise on worker performance and perception. Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, lwang4@unl.edu)

This paper presents results from a number of studies that investigated the effects of noise from building mechanical systems on human task performance and perception. Three phases of research were conducted, each of which utilized a different set of noise signals produced by building mechanical systems: (1) broad-band noise at different levels and spectral qualities, (2) tonal noise conditions, and (3) noise conditions with time-varying fluctuations. In each phase, six different noise signals (many based on *in-situ* measurements) were reproduced in an office-like setting. Thirty participants completed tasks (e.g., typing, grammatical reasoning, and math) plus subjective questionnaires, while exposed to each noise condition for up to 1 h. In general, no statistically significant differences were found in task performance across the various noise signals tested. However, higher annoyance/distraction responses from the test subjects were often significantly correlated with reduced typing performance. These higher annoyance/distraction responses were also closely correlated to higher subjective ratings of loudness, followed by roar, rumble, and tones or fluctuations. Of particular interest is that a greater perception of low-frequency rumble was significantly linked to reduced performance on both the routine and cognitively demanding tasks. [Work supported by the American Society of Heating, Refrigeration and Air-Conditioning Engineers.]

8:45—8:50 Questions

8:50—9:00 Announcements

Session 3aNCb

NOISE-CON, Noise, and Structural Acoustics and Vibration: Aircraft Interior Noise

Vincent Cotoni, Chair
ESI, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130

Contributed Papers

9:15

3aNCb1. Sound transmission through double wall aircraft structures/ mounting effect. Neple Pascale (Airbus Operations SAS, 316 Rte. de Bayonne, 31060 Toulouse Cedex 09, France, pascale.neple@airbus.com), Atalla Noureddine, and Bolduc Maxime (GAUS, Univ. of Sherbrooke,)

In the context of noise control treatment identification for turbulent boundary layer excitation, the main objective of this work is to assess the impact of mechanical links on the sound transmission loss through a double wall aircraft structure. An assembly of a stiffened panel made of composite fiber reinforced plastic (with a critical frequency around 6000 Hz) and a trim panel with a honeycomb core (with a critical frequency around 3000 Hz) has been considered in the whole study. The transmission loss of this double panel filled with glass wool blankets has been measured under diffuse sound field and point force excitation in the 100–10000 Hz frequency range for a fully decoupled configuration and a totally coupled one (rigid mounting). Extensive simulation using transfer matrix methodology has been conducted on the double panel for different glass wool filling rates, density, sound barriers, and add-on damping patches, for fully decoupled and rigid configuration. The main result is that the rigid mounting decreases significantly (up to 10 dB at 1000 Hz) the double wall transmission loss compared to a fully decoupled mounting and the acoustic efficiency of the passive treatments as well, whatever the excitation field.

9:30

3aNCb2. Validation of a polyimide foam model for use in transmission loss applications. J. Stuart Bolton, Kwanwoo Hong (Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47907-2031, bolton@purdue.edu), Tongan Wang, John Maxon (Gulfstream Aerosp. Corp., Savannah, GA 31402-2206), Brian Howerton (Lockheed Martin Exploration & Sci., Hampton, VA 23681-2199), Roberto J. Cano, Erik S. Weiser (NASA Langley Res. Ctr., Hampton, VA 23681-2199), Tyler Lorenzi (NASA Langley Aerosp. Res. Summer Scholars Program, Hampton, VA 23681-2199), Brian J. Jensen, and Richard J. Silcox (NASA Langley Res. Ctr., Hampton, VA 23681-2199)

In this paper, the use of polyimide foam as a lining in double panel applications is considered. Polyimide foam has a number of attractive functional attributes, not the least of which is its high fire resistance, thus making its use desirable in some sound transmission applications. The configuration studied here consisted of two 0.04×94 thick, flat aluminum panels separated by 5 in., with a 3 in. thick layer of foam centered in that space. Random incidence transmission loss measurements were conducted on this buildup, and conventional poro-elastic models were used to predict the performance of the lining material. The Biot parameters of the foam were determined by a combination of direct measurement (for density, flow resistivity and Young's modulus) and inverse characterization procedures (for porosity, tortuosity, viscous and thermal characteristic length, Poisson's ratio, and loss factor). The inverse characterization procedure involved matching normal incidence standing wave tube measurements of absorption coefficient and transmission loss of the isolated foam with finite element predictions. When the foam parameters determined in this way were used to predict the performance of the complete double panel system, reasonable agreement between the measured transmission loss and predictions made using commercial statistical energy analysis codes was obtained.

9:45

3aNCb3. Analysis of sound transmission through periodic structures typical for aircraft fuselages. Vincent Cotoni (ESI Group, 12555 High Bluff Dr., San Diego, CA 92130, vincent.cotoni@esi-group.com)

A novel numerical method is applied to calculate radiation properties and sound transmission loss of periodic structures typical for aircraft structures. The concept rests on that an accurate statistical energy analysis representation of the structure is determined from a small finite element (FE) model consisting of one or a few cells of the periodic structure. The FE model can be kept small and computationally efficient and can for this reason be used for parametric studies of the effect of design changes. Four different structures are analysed: a flat and a curved rib-stiffened panels typical of an aircraft fuselage, and an aluminium honeycomb inner floor panel. Calculated transmission loss results are found to compare well to measured data. Also mechanical power inputs due to point force excitations as well as radiation efficiencies are calculated and compared to measured data.

10:00

3aNCb4. Vibroacoustic characterization of a commercial airliner bulkhead. Albert Allen and Mark Moeller (Spirit AeroSystems, P.O. Box 780008, MC K95-79, Wichita, KS 67278, albert.allen@spiritaero.com)

The commercial airliner bulkhead design candidate of concern is an internal divider wall consisting of an aluminum forward panel and liner aft panel sandwiching an array of closely spaced interstitial aluminum I-beams. The main purpose of the bulkhead, which separates the cargo bay from the occupied area, is to protect the crew from being compromised by cargo during a hard landing. Occupied area noise requirements also affect the bulkhead design as it is a primary path of noise transmission from the cargo bay. Vibroacoustic performance of the bulkhead analyzed with statistical energy analysis is supported by coupon testing of a representative bulkhead test section. Fibrous acoustic treatments were also evaluated experimentally for noise reduction potential. The experimental methods, results, and component level modeling complement each other by demonstrating a vibroacoustically appropriate bulkhead design and good bulkhead modeling practices.

10:15—10:30 Break

10:30

3aNCb5. A semi-analytical method to predict the vibroacoustics response of composite and isotropic stiffened panels. Abderrazak mejdi and Noureddine Atalla (Dpartement de Gnie mécanique, Univ. of Sherbrooke, 2500 Boul. de l'Universit Sherbrooke, PQ J1K 2R1, Canada, abderrazak.mejdi@usherbrooke.ca)

The vibroacoustics behavior of aircraft-type stiffened panels is classically analyzed using deterministic methods such as the finite element and boundary element method at low frequencies or energy based methods at higher frequencies. In the present work, a general semi-analytical method based on modal expansion technique is developed to predict the vibration and acoustics radiation of both metallic and composite flat stiffened panels. Both unidirectional and bidirectional stiffened panels with eccentric stiffeners and various shapes are analyzed using the same matrix formulation. The presented model is also able to predict the response of both regular and irregular stiffened panels. The contributions of the force and moment modal coupling at each beam location are accounted for together with the effect of interaction between ribs in the case of orthogonal stiffened plate. The response to various types of excitations (point force, diffuse acoustic field, and turbulent boundary layer) are presented in terms of their joint acceptance. The model is numerically validated by comparison with the FEM/BEM and hybrid FEM/SEA methods for various configurations and excitations. Excellent agreement is found.

10:45

3aNCb6. Quiet honeycomb panel. Dan Palumbo and Jake Klos (NASA Langley Res. Ctr., MS 463, Hampton, VA 23681, d.l.palumbo@nasa.gov)

Sandwich honeycomb composite panels are lightweight and strong and, therefore, provide a reasonable alternative to the aluminum ring frame /stringer architecture currently used for most aircraft airframes. The drawback to honeycomb panels is that they radiate noise into the aircraft cabin very efficiently provoking the need for additional sound treatment which adds weight and reduces the material's cost advantage. A series of honeycomb panels incorporating different strategies aimed at reducing the honeycomb panels' radiation efficiency while at the same time maintaining its strength. The majority of the designs were centered around the concept of creating areas of reduced stiffness in the panel by adding voids and recesses to the core. The effort culminated with a reinforced/recessed panel which had 6 dB higher transmission loss than the baseline solid core panel while maintaining comparable strength. Attempts were made to damp the panels' vibration energy by the addition of lightweight particles to the honeycomb cells. These designs were very difficult to build given the particles' tendency to pollute the bond interface between the honeycomb and the face sheet. Well constructed panels exhibited very little benefit from the treatment that could not be attributed to the added mass alone.

11:00

3aNCb7. Prediction of the cabin noise of business jet using statistical energy analysis. Washington de Lima and Apoorv Ravindran (Cessna Aircraft Co., One Cessna Bldv., M.S. C3-353, Wichita, KS 67215-1424, washdelima@hotmail.com)

The application of deterministic method (FEM and BEM) to predict vibroacoustic response of complex system like an aircraft fuselage is limited to low frequency due to large size of computational model for high frequency and the high frequencies modes are very sensitive to variation in the system properties which introduces substantial uncertainty. Energy based method such as energy statistical analysis (SEA) is an alternative that avoid these difficulties. This paper presents the vibroacoustic response of business jet using SEA. The SEA model is validated using in-flight tests with the aircraft in various acoustic package configurations.

11:15

3aNCb8. Modeling freighter aircraft interior number. Mark Moeller and Albert Allen (Spirit AeroSystems Inc., Mark J. Moeller Spirit AeroSystems, Inc., M.S. K95-79, P.O. Box 780008, Wichita, KS 67278, mark.j.moeller @spiritaero.com)

A freighter version of a commercial passenger airplane platform was developed. The new architecture includes an occupied space immediately aft of the flight deck and forward of the cargo space. A bulkhead separating the occupied space from the cargo space is designed to protect people from being compromised by cargo during a hard landing. Occupied space noise requirements affect the bulkhead design as it is a primary path of noise transmission from the cargo bay. Statistical energy analysis (SEA) noise models were developed to evaluate the noise performance of the bulkhead. Models were developed for ground test, flight test, and component configurations, respectively. The component model is described in a companion paper. A fourth, requirements model, was developed to demonstrate bulkhead noise compliance. The model development and correlation efforts are detailed in this document. SEA model results are shown to correlate reasonably well with the test data.

11:30

3aNCb9. Impact of damping uncertainty on sea model response variance. Noah Schiller, Randolph Cabell (NASA Langley Res. Ctr., M.S. 463, Hampton, VA 23681, noah.h.schiller@nasa.gov), and Ferdinand Grosveld (Lockheed Martin Exploration and Sci.)

Statistical energy analysis is commonly used to predict high-frequency vibroacoustic levels. This statistical approach provides the mean response over an ensemble of random subsystems that share the same gross system properties such as density, size, and damping. Recently techniques have been developed to predict the ensemble variance as well as the mean response. However these techniques do not account for uncertainty in the system properties. In the present paper uncertainty in the damping loss factor is propagated through the analysis to obtain more realistic prediction bounds that account for both ensemble and damping variance. The analysis is performed on a stiffened composite cylinder representative of an aircraft fuselage. Realistic bounds on the damping loss factor estimates are determined using measurements acquired on the stiffened composite cylinder in the lab. The analysis demonstrates that uncertainty in system properties, such as damping, should be considered when computing the variance of the response.

11:45

3aNCb10. Hybrid characterization of engine isolator and its application to cabin noise studies. Marshall Downing (Lord Corp., 2000 W Grandview Blvd., Erie, PA 16509, mark_downing@lord.com)

Aircraft companies typically build a global acoustic model of the aircraft, engines, and engine attachments, which would allow for prediction of cabin noise related to engine operating conditions up to 500 Hz. The objective of the hybrid characterization method is to provide a matrix of stiffness (or mobility) for an isolator over a broad frequency range which is suited for use in the global model. The method is called "hybrid" because it aims at aligning viscoelastic material test data, isolator shaker testing, and a detailed finite element model. The result is a six degrees of freedom matrix of translational, rotational, and cross coupling terms for the engine attachment. This paper covers a case study showing how this method was used for the engine attachments on a business jet aircraft.

WEDNESDAY MORNING, 21 APRIL 2010 GRAND BALLROOM III/IV, 9:15 TO 11:45 A.M.

Session 3aNCc

NOISE-CON and Noise: Noise and Vibration in the Mining Industry

David Yantek, Chair
NIOSH, 626 Cochrans Mill Rd., Pittsburgh, PA 15236

Contributed Papers

9:15

3aNCc1. From development to evaluating effectiveness in industry: Building a research model for noise control technology efforts. Dana Reinke and Adam Smith (NIOSH, 626 Cochrans Mill Rd., P.O. Box 18070, Pittsburgh, PA 15236, dreinke@cdc.gov)

The NIOSH mining hearing loss research program recognizes the importance of not only developing effective noise controls but also evaluating the technology transfer and industry usage of control technologies. As a way to focus on the direct impact of NIOSH-developed control technologies on reducing noise induced hearing loss, we have developed a structured research model with metrics that help to more accurately evaluate and measure the effectiveness of noise control technologies. Through the development of a process model, we are creating a more structured process to analyze the effectiveness of noise control products through development, refinement, promotion, and long term evaluation stages. With metrics identified at each stage, the model creates a more transparent and consistent process to evaluate control technologies. This presentation describes the architecture of the process model and its associated metrics to reduce occupational noise-induced hearing loss in the mining industry. Proven control technologies for roof bolter and continuous mining machinery will serve as examples of where current research gaps occur and how they are being addressed with this new model. This research structure for noise control development, dissemination, and evaluation of control technology effectiveness can serve as a model for noise control researchers and academics.

9:30

3aNCc2. Acoustic assessment of pneumatic and electric jackleg drills used in the mining industry. Hugo Camargo, J. Shawn Peterson, and Peter Kovalchik (NIOSH, 626 Cochrans Mill Rd., P.O. Box 18070, Pittsburgh, PA 15236, hcamargo@cdc.gov)

Hearing loss is the most prevalent disease among miners. A study of U.S. western hard-rock mines revealed that 96% of machine operators were overexposed to noise, with jackleg drill operators having the most rapid noise dose accumulation. Traditionally, jackleg drills have been driven by pneumatic power. However, there are currently available rotary hammer drills powered by electricity. This paper presents an acoustic assessment of pneumatic and electric jackleg drills that involved noise source identification (NSID), penetration rate and operator's cumulative noise dose measurements, and determination of sound power levels. NSID using microphone phased array technology revealed two dominant noise sources for the electric drill, one located at the drill and one located at the drill steel-rock interaction place. In contrast, NSID for the pneumatic drill showed only one dominant noise source located at the drill. Penetration rate and noise dose measurements were combined to estimate the accumulated dose and time required to drill a reference depth hole. Sound power level measurements

yielded overall 115 and 125 dB(A) for the electric and pneumatic drills, respectively. The results show that the acoustic performance of the electric drill, despite its slower penetration rates, overcomes the benefits of traditional pneumatic drills.

9:45

3aNCc3. A control suite to reduce roof bolting machine drilling noise. M. Jenae Lowe, J. Shawn Peterson, Lynn Alcorn, and David Yantek (Natl. Inst. for Occupational Safety and Health, 626 Cochrans Mill Rd., P.O. Box 18070, Pittsburgh, PA 15236, jli7@cdc.gov)

Among underground coal miners, hearing loss remains one of the most common occupational illnesses. In response, the National Institute for Occupational Safety and Health (NIOSH) conducts research to reduce the noise emission of various underground coal mining equipment, an example of which is a roof bolting machine. Field studies support the premise that, on average, drilling noise is the loudest noise that a roof bolting machine operator would be exposed to and contributes significantly to the operators' noise exposure. NIOSH has determined that the drill steel radiates a significant amount of noise during drilling. NIOSH in collaboration with Corry Rubber Corporation has developed a suite of controls to reduce drilling noise which consists of a bit isolator, chuck isolator, and a Collabsible Drill Steel Enclosure. This control suite effectively reduces the noise radiated by the drill steel. Laboratory testing confirms that the control suite reduces sound pressure levels generated during drilling by 13 dB(A) at the operator's location. As a result of this reduction in drilling noise, the noise exposure of the roof bolting machine operator is reduced.

10:00

3aNCc4. The development of a damped drill steel to reduce roof bolting machine drilling noise. J. Shawn Peterson, Hugo Camargo (Natl. Inst. for Occupational Safety and Health, P.O. Box 18070, Cochrans Mill Rd. Pittsburgh, PA 15236, JPeterson@cdc.gov), and Greg Goetchius (Mater. Sci. Corp.)

Among underground coal miners, hearing loss remains one of the most common occupational illnesses. In response, the National Institute for Occupational Safety and Health (NIOSH) conducts research to reduce the noise emission of various underground coal mining equipment, an example of which is a roof bolting machine. Field studies support the premise that, on average, drilling noise is the loudest noise that a roof bolting machine operator would be exposed to and contributes significantly to the operators' noise exposure. NIOSH has determined that the drill steel radiates a significant amount of noise during drilling. NIOSH has conducted research to evaluate a damped drill steel to reduce noise radiation. With this reduction in drilling noise, the noise exposure of the roof bolting machine operator is reduced. This paper documents the research to date of this damped drill steel. Laboratory testing confirms that the concept of a damped drilll steel reduces sound levels generated during drilling, and by extension, operator noise dose exposure.

10:15—10:30 Break

10:30

3aNCc5. Drill rod Coulomb friction damper. Anne Guthrie and Henry Scarton (Dept. of Mech., Aerosp. and Nuclear Eng., Rensselaer Polytechnic Inst., Troy, NY 12180-3590, scarton@rpi.edu)

The acoustical performance of a case-hardened Coulomb friction damper consisting of a case-hardened roll-threaded collar resting against a reverse taper segment of drill rod is discussed. Because of the axial impact, many transverse bending modes of vibration are excited in long prismatic drill rod, along with longitudinal and torsional resonances. The self-tightening nature of the collar that is caused to simultaneously move axially due to its inertia and rotate on the rolled thread and consequently press on the case-hardened tapered surface of the drill rod induces interfacial frictional rubbing which is conclusively shown to be present resulting in 30 dB of attenuation of the modes of vibration. These observations are experimentally shown and theoretically verified.

10:45

3aNCc6. Evaluation of stiffeners for reducing noise from horizontal vibrating screens. David Yantek (CDC-NIOSH-PRL, 626 Cochran Mill Rd., Pittsburgh, PA 15236, dyantek@cdc.gov) and Shawn Catlin (Mallett Technol., Inc.)

Sound levels around vibrating screens in coal preparation plants often exceed 90 dB(A). The National Institute for Occupational Safety and Health (NIOSH) is developing noise controls to reduce noise generated by horizontal vibrating screens. NIOSH determined the vibration mechanism housings and the screen body to be the dominant noise sources on a horizontal vibrating screen. Researchers used beamforming, experimental modal analysis, and operating deflection shape analysis to examine noise radiated by the screen body. Based on these results, finite element analysis and a NIOSH-written program were used to estimate the sound power level reduction resulting from adding rib stiffeners to key locations on the screen sides. Rib stiffeners made from two sizes of steel channel and two different cross-sections, C and T, were evaluated. In addition, the effects of orienting the stiffeners horizontally and vertically were examined. Finally, the stiffeners were evaluated with the ends welded to the existing ribs on the screen sides and with the ends free. The results indicate that for a broadband input, the smaller T cross-section oriented vertically with the ends welded to the existing ribs was the best option. This configuration reduced the predicted *A*-weighted sound power level by 7 dB.

11:00

3aNCc7. An investigation of the noise dynamics in a southern illinois underground coal mine. Marek Szary, Yoginder Chugh, William Bell (Carbondale College of Eng., Southern Illinois Univ., 1230 Lincoln Dr., Carbondale, IL 62901-6603, szary@engr.siu.edu), and Joseph Hirschi (Illinois Clean Coal Inst.)

Noise in an underground coal mine has dominant components generated mainly from three sources: (1) continuous mining machine (CMM), (2) roof bolters, and (3) cars/vehicles which are transporting personnel end/or coal. Each of these three noise sources also has a number of well defined subsources with their own noise characteristics. The CMM noise is comprised mainly of noise generated by coal cutting drum, wet scrubber for dust control, and coal transport conveyor (called also the CMM's tail). Roof bolter's noise is generated during the drilling of the roof bolt holes in the bolting process. Personnel and coal transportation vehicles generate noise from the power driven system. The personnel most exposed to these noises are operators of these machines and associated support personnel. Three selected techniques with appropriate instrumentation were used to monitor exposure of the personnel to the noise and noise energy over a period of time. The most common technique is based on the use of personal noise dosimeters. The sound level meters (both pressure and power) were also used to collect noise data in form of instantaneous readings and also to check calibration of other sound measuring instruments. Most useful information was obtained from continuous recordings of the noise over time. This paper discusses the variability or dynamics of the generated noise in both frequency and time domains

11:15

3aNCc8. Evaluation of noise exposures of miners operating highwall mining equipment. Link Bowers (U.S. Dept. of Labor, Mine Safety and Health Administration, P.O. Box 18233, Cochrans Mill Rd., Bldg. 038, Pittsburgh, PA 15236, bowers.link@dol.gov)

Modern highwall mining systems are starting to replace surface coal augers in some highwall mining operations because of their ability to increase coal production and recovery rates. Another benefit of using highwall mining systems instead of surface coal augers is the reduction in noise exposures to the affected miners extracting the coal from the highwall. Several surface coal augers and highwall mining systems were evaluated for full shift noise exposures of the machine operator and helpers during this investigation. Engineering noise controls were also evaluated for each piece of highwall mining equipment. The operator and helper working on a modern highwall mining system were exposed to average sound pressure levels

of 61.5 and 89.0 dBA, respectively, while the operator and helper working on a surface coal auger were exposed to higher average sound pressure levels of 95.4 and 94.0 dBA, respectively.

11:30

3aNCc9. Wave propagation characteristics in mud-filled drilling pipe. Abhay Patil and Yong-Joe Kim (Dept. of Mech. Eng., Texas A&M Univ., College Station, TX 77843, abhyapatil@neo.tamu.edu)

The objective of this study is to analytically and experimentally analyze the wave propagation characteristics of a mud-filled drilling pipe in terms of wave speed and spatial decay rate. Here, the pipe is modeled as an elastic hollow pipe and the inside mud as a fluid. A wave equation of the mud is then derived based on the assumption of one-dimensional plane wave propagation that is valid in low frequencies. This wave equation is coupled with the equation of motion of the elastic pipe. For the purpose of validating the analytical wave solutions of the wave equation, an experiment is performed with a mud-filled pipe. The pipe is driven by a piston at one end and the resulting dynamic responses are then measured with an array of strain gauges installed on the external surface of the pipe along the axial direction at the other end. A wave decomposition procedure is applied to the measured data to separate direct and reflective waves. The direct waves are then used to calculate their phase speeds and spatial decay rates. It is shown that the experimentally measured wave propagation characteristics are successfully used to validate and improve the analytical model.

Session 3aNCd

NOISE-CON and Noise: Ground-Borne Noise and Vibration Modeling

James T. Nelson, Chair
Wilson, Ihrig Associates, Inc., 5776 Broadway, Oakland, CA 94618

Contributed Papers

9:15

3aNCd1. Deep dynamic compaction vibrations impacts on neighboring structures. Amir Yazdanniyaz and Sen Bui (Acoust. Eng. Services, Inc., 22801 Crespi St., Woodland Hills, CA 91364, amir.yazdanniyaz @aesacoustics.com)

This paper describes results from vibration study conducted for a construction of proposed building structures on a covered landfill site. Construction site is surrounded by various land uses, including mobile homes, single dwelling units, and infrastructures. Deep dynamic compaction (DDC) method is utilized to provide soil compactions necessary for the construction of the buildings and parking lot foundations. Concern with the potential structural damage and the human perceptions of the ground vibration exist because of close proximity of the existing structures (i.e., mobile homes, single homes, roadways, and bridges) to the compaction area. DDC induced ground-vibrations level were modeled using actual vibration measurements "pilot tests" and the site's soil dynamic characteristics. Vibration mitigation plan were developed to minimize the impacts.

9:30

3aNCd2. Vibration impact of 150 MW cogeneration power plant on a semiconductor fabrication facility. Ahmad Bayat and Jon Byron Davis (490 Post St., Ste. 1427, San Francisco, CA 94102, abayat@vaconsult.com)

A 150 MW cogeneration power plant construction was planned next to an extremely vibration-sensitive semiconductor fabrication facility in Taiwan. Both facilities were under construction simultaneously. We were asked to evaluate the impact of the power plant vibration sources on the semiconductor facility and determine that they can coexist at only about 30 m separation distance. Some of the power plant sources included large rotating equipment such as 7000-HP heat extraction blowers, turbines, and pumps. Dynamic forcing functions were extracted from a similar plant. *In situ* transfer function mobilities were obtained using seismic load drop equipment. Predictions were made which were in-line with the as-built results with the exception of two blowers. Their foundation design had to be retrofitted with deep pile foundations to transfer their vibration to deeper stiffer soil.

9:45

3aNCd3. Techniques for measuring the propagation speed of vibration waves in the ground. Anthony Nash (Charles M. Salter Assoc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, anthony.nash@cmsalter.com)

Characterizing the transmission of ground vibration is important for assessing impacts upon buildings located adjacent to a rail right-of-way. This type ground-borne vibration transmission is typically dominated by surface or "Rayleigh" waves. When predicting the impacts, it is helpful to understand the attenuation of vibration from one location to another along the surface of the ground. In many cases, measuring the time-averaged rms vibration at spaced locations away from a line source is sufficient if one wishes only to quantify the reduction in ground borne energy density. For other purposes, however, it is essential to know the speed (and length) of the propagating wave front. This paper discusses successful (and not so successful) techniques for measuring the propagation speed of Rayleigh waves in the ground using both steady state and transient "point" sources. In terms of meaningful data, the steady state vibration method was found to be inferior to the transient vibration method. The steady state technique generated contradictory information on phase and amplitude at the spaced measurement locations. In contrast, the transient technique always indicated that (1) the Rayleigh wave propagated outward from the source and (2) the vibration amplitudes decreased between the near and far measurement locations.

10:00

3aNCd4. Modeling ground response functions in rock: A comparison of multiple methods. Gary Glickman and Richard Carman (Wilson, Ihrig & Assoc., 65 Broadway, Ste. 401, New York, NY 10006, gglickman@wiai .com)

Rail transit operations often produce groundborne vibration with energy that extends into the audible frequency range. It is often necessary to predict potential ground/structure-borne vibration and noise transmission in the design of new buildings in close proximity to existing rail lines or for new rail facilities/alignments. In a detailed analysis, key factors include source vibration characteristics (spectral force input to soil and/or track structure), ground vibration propagation characteristics (spreading and material damping losses), building foundation type (soil/structure interaction and coupling losses), and structural design of upper floors (floor and wall resonances and attenuation through the building). This paper focuses on methods for obtain-

ing vibration propagation characteristics in rock which is an efficient transmitter of vibration at audible frequencies. Three methods for obtaining one-third octave band response functions are compared: empirical methods, finite element analysis, and seismic modeling.

10:15—10:30 Break

10:30

3aNCd5. Observed frequency and depth effects on attenuation of ambient ground vibration. Hal Amick and Michael Gendreau (Colin Gordon & Assoc., 150 North Hill Dr., Ste. 15, Brisbane, CA 94005, hal.amick@colingordon.com)

This article presents observed relationships between depth and frequency with regard to ambient ground vibration. Typically, site vibration studies are conducted on undeveloped locations (grass fields, parking lots, etc.) to assess the ambient vibration conditions for a potential building site. However, the vibration-sensitive equipment of concern is frequently placed below grade, such that the final ambient vibration conditions it experiences might be quite different from that shown in the field survey. An earlier study [NoiseCon (2007)] presented data illustrating how several factors—including building foundation stiffness, ground stiffness, and building geometry—all have an effect on the propagation from "ambient" vibration sources. This presentation extends that work, specifically examining the relationship between depth, frequency, and extent of attenuation associated with below-grade spaces.

10:45

3aNCd6. Variation in measured force density levels of light rail vehicles. Shankar Rajaram and Hugh Saurenman (ATS Consulting, 801 S. Grand Ave, Ste. 575, Los Angeles, CA 90017, srajaram@atsconsulting.com)

The procedure recommended by the Federal Transit Administration for detailed predictions of ground-borne vibration is based on measurements at an existing rail system and measurements at the target site. A force density level (FDL) is derived from train vibration and line source transfer mobility (LSTM) measurements at the existing system. This FDL is then combined with the LSTM measured at the target site to develop predictions. This empirical procedure usually provides more reliable predictions than computer modeling procedures. FDL is assumed to characterize the trains and track support system and LSTM is assumed to characterize the effects of local geology. Implicit assumptions when applying this procedure are that FDL is independent of the local geology and that the vehicle and track support system will be the same at the target site as they are at the vehicle test site. This paper considers the differences in FDLs measured at light rail systems in Los Angeles, Minneapolis, Phoenix, and Portland. Issues discussed are how to adjust for train speed, why the speed adjustments are system specific, causes for FDL variations when the vehicles and track support systems are nominally the same, and procedures for obtaining consistent FDL measurements.

11:00

3aNCd7. Analysis and design of new floating slab track for special trackwork using finite element analysis. James Phillips (Wilson, Ihrig & Assoc., Inc., 5776 Broadway, Oakland, CA 94618, jphillips@wiai.com)

This paper discusses the development of a new vibration isolation system for tunnel crossovers in a future rail line extension. The previous design included ballasted track on top of continuous concrete slabs on rubber vibration isolators. The goal of the new design was to replace the ballast and tie track with rails fastened directly to discrete slabs similar to the vibration isolation system utilized on running portions of track within existing tunnels. Finite element analysis (FEA) was used to evaluate the dynamic interaction between the vehicle, rails, resilient rail fasteners and the vibration isolation system. The results of the analyses guided the design in determining the size and thickness of the inertia slabs, the stiffness and design of the rubber vibration isolation pads, and the rail fastener selection.

Session 3aNSa

Noise, Physical Acoustics, and INCE: Military Noise Environments—Continuous and Impulsive I

Richard L. McKinley, Chair
Air Force Research Lab., 2610 Seventh St., Wright Patterson Air Force Base, OH 45433-7901

Chair's Introduction—8:05

Invited Papers

8:10

3aNSa1. Hearing loss mitigation through applied acoustics. Kurt Yankaskas (Office of Naval Res. (Code 342), 875 North Randolph St., Arlington, VA 22203-1995, kurt.d.yankaskas@navy.mil)

Hearing loss in the Navy and Marine Corps has been accelerating for decades, reducing combat effectiveness and creating hazardous conditions. This paper will discuss applied acoustics in current and future programs and its related challenges. Source noise reduction limits the amount of energy transmitted to the structure or into the environment. Acoustic modeling, insulation/isolation, and personal hearing protection are vital tools in this process, but often adding weight and complexity. Source noise solutions such as isotropic superfinishing treatment of gears increases transmitted power and reduces fuel consumption and lubricant temperatures while increasing gear life and maintenance intervals. Great strides are also currently being made in the arena of personal hearing protection to reduce hearing loss rates in environments where reducing source noise is not a viable option, such as the aircraft carrier flight deck. Technological advances in insulation and isolation and possible operational changes will also be discussed, adding to the tools required to approach this issue from an integrated systems approach. In conclusion, a systems integration approach to incorporating noise reduction into the design of vessels, vehicles, and platforms from the start of the design process will reduce life cycle costs, benefiting the Navy and Marine Corps.

8:30

3aNSa2. Navy aircraft carrier near-field jet noise and flight deck crew hearing protection. James A. Janousek (Naval Air Systems Command, 48110 Shaw Rd., Bldg. 2187, Patuxent River, MD 20670, james.janousek@navy.mil), Valerie S. Bjorn (Naval Air Systems Command, Arnold AFB, TN 37389-6000), Richard L. McKinley (Air Force Res. Lab., 2610 Seventh St., Wright-Patterson AFB, OH 45433-7901), and James K. Wilt (Naval Air Systems Command, Patuxent River, MD 20670)

Military jet operations on board US Navy aircraft carriers produce excessive acoustical noise fields in which personnel work. For those working in these noise environments, hearing impairment generally starts and worsens as they continue to perform countless high-stakes jet aircraft launch and recovery missions. Several noise measuring surveys were completed to characterize near-field jet noise on US Navy flight decks. These data set the foundation for numerous efforts to improve personal hearing protection and reduce jet engine noise.

8:50

3aNSa3. Near-field noise of high-performance military aircraft and required hearing protection. Richard L. McKinley (Air Force Res. Lab., 2610 Seventh St., Wright-Paterson Air Force Base, OH 45433-7901, richard.mckinley@wpafb.af.mil)

The near-field acoustic environment of a high-performance military jet aircraft is very complex and intense. The overall sound pressure levels at high-engine power settings can exceed 170 dB very near the aircraft and can be in the 130–150 dB range at personnel locations. This paper describes the measured near-field noise environments from several Air Force and Navy high performance aircraft such as the F-14, F-15, F-16, F-18, and F-35. This paper will also describe a technique of calculating the hearing protection requirements in octave bands using a baseline exposure criterion and noise dose.

9:10

3aNSa4. Characterization of near- and far-field nonlinearity in high-performance military jet aircraft noise. Tracianne B. Neilsen, Kent L. Gee (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602, tbn@byu.edu), J. Micah Downing, and Michael M. James (Blue Ridge Res. and Consulting, Asheville, NC 28801)

This study addresses the characterization of nonlinearity in military jet aircraft noise data in both the near and far fields. Previous work [Gee *et al.*, J. Acoust. Soc. Am. **123**, 4082–4093 (2008)] showed using a generalized Burgers equation-based model that nonlinear propagation effects were significant in far-field, ground run-up data from the F-22 Raptor for different engine powers and propagation angles. In this investigation, those findings are further vetted using data collected from the F-35 Joint Strike Fighter over greater propagation ranges and angles. The nonlinear nature of the near fields of the F-22 and F-35 jet noise is characterized through the calculation of maps of the waveform and waveform time derivative skewness as well as the crest factor. These results provide insights regarding short-range shock formation and the apparent origin of the nonlinearity-producing region of the plume. [Work supported by the Air Force Research Laboratory.]

9:30

3aNSa5. Weapon noise exposure of the soldier and its measurement. Karl Buck and Pascal Hamery (ISL Group APC, 5 r. du Gal. Cassagnou, 68301 Saint-Louis, France, karl.buck@isl.eu)

On the battlefield a soldier is continuously exposed to various types of noise. Much of this exposure is noise generated by his own weapon or by weapons of close by troops. The exposure levels are between 160-dB peak for small arms and 190-dB peak at the soldier's ear for some antitank weapons, with A-durations from 0.3 ms (small caliber) to 4 ms for large caliber weapons (e.g., Howitzers). As these parameters feed into the damage risk criteria which are in use by the armed forces of different countries, it is important that the recorded pressure time histories are as precise as possible. The paper will present the acoustic levels to which a soldier may be exposed by different types of weapon and under different circumstances. It also will present measurement techniques to use for reliable results.

9:50

3aNSa6. Measurement and characterization of army impulsive noise sources. Joel T. Kalb (U.S. Army Res. Lab., Aberdeen Proving Ground, MD 21005-5425, joel.kalb@us.army.mil)

Peak pressure levels in army noise environments range from 140- to 195-dB near soldiers firing energetic weapons and are all dangerous to unprotected ears. Blast reverberation in confined spaces and steady noise in armored vehicles is also hazardous, so complete assessment requires the widest possible measurement range. To accurately measure pressures in the free-field or under hearing protectors, microphone considerations include type, location, orientation, mounting, shape, and shielding from flash or mechanical shock. Blast-waves rise instantaneously and return to free-field ambient within 0.3–15 ms, while in enclosures this can increase to 100 ms, requiring wide bandwidth signal processing. In addition to hazard, acoustic detection and annoyance at greater distances from weapons involve propagation effects in the atmosphere and at the ground which alter the wave-shape.

10:10—10:25 Break

10:25

3aNSa7. Unique challenges for hearing protection in military noise environments. Douglas S. Brungart, Mary T. Cord, and Kim L. Block (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave NW, Washington, DC, 20307, douglas.brungart@us.army.mil)

Military noise environments, and, in particular, the noise environments faced by dismounted soldiers on the battlefield, are characterized by wide variations in ambient level. Situations can quickly and unexpectedly change from quiet conditions where the sound of a snapping twig might alert the listener to hostile enemy activity to extreme noise conditions where firefights, explosions, or loud machinery create levels of noise that can cause hearing loss in a matter of minutes or seconds. This poses a unique challenge for the designers of military hearing protection, who must produce systems that provide enough comfort and acoustic transparency to convince users to consistently wear them in quiet conditions to ensure that they will be in place if an unexpected blast or other potentially damaging noise exposure occurs. In this talk, we discuss some of the issues that can influence the willingness of listeners to wear hearing protection for extended periods in quiet environments, including acoustic factors that have been shown to influence situational awareness while wearing hearing protection. Additionally, lessons on the factors impacting the comfort of hearing devices drawn from studies of how consistently hearing aids are worn by hearing impaired listeners will be discussed.

Contributed Papers

10:45

3aNSa8. Improved system for detection, localization, and classification of military impulse noise. Jeffrey Allanach, Justin Borodinsky (Appl. Physical Sci. Corp., 475 Bridge St., Ste. 100, Groton, CT 06340, jallanach@aphysci.com), Jeffrey Vipperman, and Matthew Rhudy (Univ. of Pittsburgh, Pittsburgh, PA 15261)

Impulse noise monitoring systems are needed for quantifying the magnitude and time of impulsive noise events to ensure compliance and to provide an archival record of noise emanating from military installations. Currently, there are several commercial systems; however, they report an overwhelming number of false events generated by windborne noise and distant non-military acoustic events. This can bias noise statistics to the point where a meaningful assessment of the acoustic sound levels from a site is not possible. APS, in collaboration with the University of Pittsburgh, have developed an improved noise monitoring system for mitigating windborne and other sources of non-military noise. This system includes a collection of remote sensors capable of detection, localization, and classification. Each sensor uses an acoustic array and real-time signal processing codes to estimate the noise source location. Mitigation of windborne events is accomplished using cross-channel correlation analysis, beamforming, and classification. A prototype of this system is installed at the Marine Corps Base Camp Lejeune where it has been automatically reporting and archiving detected impulse noise. In this paper, we will present the specialized array processing algorithms, system design, and its performance as determined by months of detection statistics accumulated in the field.

11:00

3aNSa9. Aircraft jet source noise measurements of a Lockheed Martin F-22 fighter jet using a prototype near-field acoustical holography measurement system. Michael M. James (Blue Ridge Res. and Consulting, 13 1/2 W. Walnut St., Asheville, NC 28801, michael.james@blueridgeres.com), Kent L. Gee, Alan T. Wall (Brigham Young Univ., Provo, UT 84602), J. Micah Downing, Kevin A. Bradley (Blue Ridge Res. and Consulting, Asheville, NC 28801), and Sally Anne McInerny (Univ. of Alabama at Birmingham, Birmingham, AL 35205)

Military jet aircraft generate high levels of noise which require innovative measurement and analysis methods to characterize the jet noise. A near-field acoustic holography (NAH) system is being developed to provide model refinement and benchmarking, evaluate performance of noise control devices, and predict ground maintenance personnel and community noise exposure. A prototype NAH system was used to perform jet source noise measurements of an F-22 at Holloman AFB, NM. This NAH system utilizes a patch and scan measurement approach using a 90 channel microphone array and over 50 reference microphones. The microphone array was used to make pressure measurements on a planar surface extending from the jet nozzle to 75 ft along the plume and over 6 ft in height. These measurements were made parallel either to the shear layer or to the nozzle center line of the jet plume at three different offset distances. Measurements were made every 6 in. on the surface which results in over 1800 measurements per offset distance. The ground based fixed reference microphones were located parallel to the nozzle center line. The measurement approach and pressure mea-

surements for four power conditions detailing the source size, frequency content, and ground effect will be detailed. [Work supported by AFRL SBIR.]

11:15

3aNSa10. Application of near-field acoustical holography to high-performance jet aircraft noise. Alan T. Wall (371 E 720 S, Orem, UT 84058), Kent L. Gee (Brigham Young Univ., Provo, UT 84602), Michael M. James (Blue Ridge Res. and Consulting, Asheville, NC 28801), and Michael D. Gardner (Graduate Program in Acoust., Penn State Univ., State College, PA 16802)

Near-field acoustical holography (NAH) measurements were made to visualize the sound field in the jet exhaust region of an F-22 Raptor. This is one of the largest-scale applications of NAH since its development in the 1980s. A scan-based holographic measurement was made using a 90-microphone array with 15-cm regular grid spacing for four engine power settings. The array was scanned through 93 measurement positions along three different measurement planes in a region near 7 m from the jet centerline and 23 m downstream. In addition, 50 fixed reference microphones were placed along the ground 11.6 m from the jet centerline, spanning 30.8 m. The reference microphones have been used to perform virtual coherence on the measurement planes. Statistically optimized NAH has been used to backpropagate the sound field to the source region for low frequencies and to identify jet noise characteristics. Ground reflection interference and other non-ideal measurement conditions must be dealt with. Details relating to jet coherence-lengths and their relation to reference microphone requirements will be discussed. Preliminary results of this ongoing work will be presented. [Work supported by Air Force SBIR.]

11:30

3aNSa11. Experiments on the directivity of sound emission from firearms. Morten Huseby and Haakon Fykse (Norwegian Defence Res. Establishment, Pb. 25, 2027 Kjeller, Norway)

Personnel firing weapons will be exposed to noise created by the expanding gunpowder gas escaping from the muzzle. The noise from a rifle is directive, up to 20 dB louder in the shooting direction than behind the weapon. Usually personnel firing a weapon is taking advantage of the lower noise level behind the weapon. However, today firearms will be used by soldiers in complex scenarios where the head of one soldier might be directly to the side or even slightly in front of the muzzle of another soldier's weapon. Most countries training regulations have not been updated to include all effects of the directivity of firearms in complex scenarios. Unexpected results were obtained from previous measurements on weapons with flash suppressors attached to the end of the barrel. The noise may be considerably larger in front of the weapon with flash suppressor fitted. The angular resolution of previous experimental data was not high enough to provide an explanation. This motivated us to design an experimental setup where the pressure wave from the weapon is measured, at 80 cm from the muzzle, in 180 directions relative to the firing direction. The measurements are done with and without a flash suppressor mounted.

11:45

3aNSa12. A hearing protector model for predicting impulsive noise hazard. Joel Kalb (Human Res. & Eng. Directorate, U.S.Army Res. Lab, Aberdeen Proving Ground, MD 21005-5425, jkalb@arl.army.mil)

Reduction in impulsive noise hearing hazard by earplugs and earmuffs is calculated with an electroacoustic lumped-parameter model of HPD attenuation using real ear attenuation at threshold (REAT) data. Energy flows into the occluded volume along three paths, each considered as a piston: (1) the rigid protector mass moving against the skin, (2) leakage at the support, and (3) transmission through the protector material (a second piston within the rigid piston). Circuit elements are adjusted so loss matches REAT data assuming path 1 is important at low, 2 at middle, and 3 at high REAT frequencies. Applying the model to 384 REAT data sets for ANSI S12.6 method B naive users gives statistical frequency distributions of occluded volume and leakage elements. For a given free-field impulsive noise, the model pressure predictions under the protector are compared to measurements on human and acoustical manikin ears to check validity of assumptions. The hearing hazards of the measured waveforms and the predicted waveforms are calculated with our previously developed auditory hazard analysis algorithm for humans ear model. The result is a cumulative frequency distribution of hazard based on user fit data useful in finding the best protector for a given impulsive noise.

Session 3aNSb

Noise, Architectural Acoustics, INCE, and ASA Committee on Standards: Standards in Psychoacoustics

Klaus Genuit, Chair
HEAD acoustics GmbH, Ebertstrasse 30a, Herzogenrath, 52134, Germany

Invited Papers

8:00

3aNSb1. Need for standardization of psychoacoustics. Klaus Genuit (HEAD Acoust. GmbH, Ebertstr. 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

In the last decades, psychoacoustic parameters gained more and more importance in the context of sound perception and assessment in various applications. In particular, loudness was introduced as a better parameter than *A*-weighted level, because it shows a much better correspondence with the subjective impression of "volume" (loudness). This parameter was the object of standardization in the last years and few loudness standards were established (DIN 45631/A1, ISO 532, ANSI S3.4). The calculation of sharpness is also standardized (DIN 45692) and the standardization of roughness is recently under consideration by the DIN. These efforts have led to a global acceptance and a widespread use of psychoacoustic parameters in the automotive and IT-field. However, since the human signal processing is complex and the development of valid calculation models difficult, psychoacoustic phenomena must further be studied. Moreover, the validity, accuracy, and applicability of psychoacoustic parameters to specific noises must be determined. Especially, in the context of the description and classification of environmental noise, the potential of psychoacoustics has not been fully used so far. This paper will show the current status of standardization of psychoacoustics and will discuss the needs for the future.

8:20

3aNSb2. Comparative study of a proposed time varying loudness standard for common sounds. Colin Novak, Helen Ule (Dept. of Mech., Automotive and Mater. Eng., Univ. of Windsor, 401 Sunset Ave., Windsor, ON N9B 3P4, Canada, novak1@uwindsor.ca), and Tomasz Letowski (Army Res. Lab. Human Res. and Eng. Directorate)

Several calculation procedures exist and are standardized for the determination of loudness of steady sounds. While loudness models for unsteady sounds do presently exist, they differ in their outcomes and have not been yet sufficiently validated using round robin or other appropriate evaluation methodology. Recently, the Deutsches Institut fur Normung (DIN) announced that it will soon issue a new standard describing a method for calculation of loudness of time-varying sounds. However, the DIN 45631-A1 takes a less common approach in that it does not fully describe a methodology to calculate the unsteady loudness. It instead offers a more general procedure along with open-ended checks and balances for verifying the progress through the various calculation steps. The caveat of this is that while conditions specified by this standard may be satisfied, varying results may be achieved due to the procedures' open architecture. Conversely, the checks and balances may be enough such that identical results are achieved for the same sounds. This study investigates several commercial applications using this standard and compares their results for several common inputs. The results of the study will be discussed in terms of the extent to which compliance with DIN 45631-A1 results in identical loudness scores.

8:40

3aNSb3. An overview of issues concerning loudness measurement versus time. Wade R. Bray (HEAD Acoust., Inc., 6964 Kensington Rd., Brighton, MI 48116, wbray@headacoustics.com)

An overview of present methodologies for loudness measurement, especially considering time-varying sounds, is given. An ideal loudness-measuring method would mimic both the spectral and temporal acuities of human hearing including the relationship of those acuities at all audible frequencies. Due in part to practical considerations in their development history, none of the present set of loudness-measurement tools offers fully accurate representation of subjective time-frequency impressions. The most general limitations are insufficient spectral resolution and frequency-dependent time resolution. A recently standardized method well matches the temporal behavior of subjective loudness perception but only above a certain frequency due to lower-frequency time-resolution limits imposed by prior standards necessitating compliance. Examples will be shown of trade-offs and varying balances among several critical factors: time and frequency resolution at all audible frequencies, representation of temporal masking and growth/release effects, and appropriate spectral representations over varying combinations of broadband and narrowband signals.

9:00

3aNSb4. Loudness models applied to technical sounds. Roland Sottek (HEAD Acoust. GmbH, Ebertstr. 30a, 52134 Herzogenrath, Germany, roland.sottek@head-acoustics.de)

Loudness evaluation has become a central focus for assuring better consideration of subjective sound intensity phenomena than frequency-weighted levels such as dB(A). Different standards are available for evaluating loudness of stationary sounds (ANSI, ISO, and DIN). In principle, standards are very helpful in daily life, even if the method does not consider the most recent research results. But a variety of standards generating different results, at least for the most important technical or industrial sounds, may lead to confusion. Considering that most of these sounds are time variant, a single model of time-varying loudness is preferable. At present the recently

published extension of the German standard for time-varying loudness (DIN 45631/A1) is the sole standardized method, applicable also for stationary sounds. There are other models allowing for loudness evaluation of time-variant sounds. Some years ago, a Hearing Model was borne out of a widespread interest to have a single standardized psychoacoustic model describing and explaining many phenomena at once. The Hearing Model is based on the physiology of human hearing and has been validated by testing against previously conducted psychoacoustic research results. The different methods will be compared and validated with subjective tests using technical sounds.

Contributed Papers

9:20

3aNSb5. Safe lifetime occupational exposure-1 LONE (lifetime occupational noise exposure). Robert D. Bruce, Arno S. Bommer, Kimberly A. Lefkowitz, and Noel W. Hart (CSTI Acoust., 16155 Park Row, Ste. 150, Houston, TX 77084)

For many years, it has been recognized by those working in the field of industrial noise that understanding how much noise is needed to cause hearing loss over a lifetime is difficult to communicate to most people, even those who have an understanding of logarithms. The concept of expressing noise exposure in industrial situations without decibels is the focus of this paper. Eldred ("Sound Exposure without Decibels" Internoise-86) discussed this approach for community noise. ANSI Standard S3.44-1996 defines sound exposure with units of Pascals squared seconds, or PASQUES, as noted by Eldred. This paper proposes that a safe value for lifetime occupational exposure to noise be expressed in terms of PASQUES. The authors will discuss the pros and cons of such an approach and offer 11.5×10^6 million PASQUES as the upper limit for a safe lifetime exposure to occupational noise.

9:35

3aNSb6. Earmuff noise attenuation by microphone in real ear technique. Samir N. Y. Gerges (Mech. Eng., Federal Universit of Santa Catarina, Florianopolis, Sc, Brazil, samir@emc.ufsc.br), Danilo A. Agurto, Jorge Arenas (Universidad Austral de Chile, Chile), Jose Bento Coelho, and Miguel M. Neves (IST, Lisbon, Portugal)

When using a hearing protector at the workplace, it is necessary to quantify its noise attenuation by laboratory measurements. This paper aims at developing an objective methodology for measuring hearing protector noise attenuation of earmuff type using "microphone in real ear" (MIRE) method. The methodology uses the "insertion loss" (IL) as parameter, which is the difference between the sound pressure level with and without the hearing protector, using MIRE, and then calculating the hearing protector IL. The results for four different hearing protectors are compared with the subjective method "real-ear attenuation at threshold" (REAT). Factors such as physiological noise and bone conduction are quantified using the relation between MIRE and REAT method. Results are compared with values obtained from the literature.

WEDNESDAY MORNING, 21 APRIL 2010 HERON, 10:10 A.M. TO 12:00 NOON

Session 3aNSc

Noise, Architectural Acoustics, and INCE: Low Frequency Noise and Ultrasonics

Detlef Krahé, Chair
Univ. of Wuppertal, Rainer-Gruenter-Str. 21, Wuppertal, 42119, Germany

Chair's Introduction—10:10

Contributed Papers

10:15

3aNSc1. New options in microphones and instrumentation for making field infrasonic measurements. Kurt Veggeberg (Natl. Instruments, 11500 N. MOPAC C, Austin, TX 78759, kurt.veggeberg@ni.com)

There are many interesting man-made, animal, and other naturally occurring infrasonic signals. Typical measurement microphones and instrumentation systems are usually ac coupled and roll-off at various frequencies below 20 Hz since they were primarily designed for audible sound. Making infrasonic measurements has often used a variety of sensors or microphones that are not uniform or compliant with international standard IEC 1094 for dimensions of measurement microphones but offer capabilities for measuring to frequency ranges below 1 Hz. In addition, they typically are not compliant with ANSI or IEC standards for sound level meters. This is an overview of new IEPE/ICP condenser type measurement microphones and instrumentation capable of making measurements below 0.5 Hz and how they can be deployed in making precision acoustic field measurements compliant with acoustic standards at very low-frequency ranges.

10:30

3aNSc2. Laboratory measurements of sound transmission at low frequencies. Chris Fuller (Dept. of Mech. Eng., VaTech, 133 Durham Hall, Blacksburg, VA 24061), Kathleen Kondylas (NEVA Assoc., Newburyport, MA 01950), and Natalia Levit (DuPont, Richmond, VA 23234, natalia.v.levit @usa.dupont.com)

The ability of acoustic materials to attenuate sound is usually determined in accordance with ASTM E-90 or ISO 140-3, both based on the diffuse field theory. However, the reality is that the size of typical reverberant chambers does not provide a sufficiently diffuse field at low frequencies. As a result, there is significant variation in sound transmission data of tested materials below 150 Hz. Several attempts were undertaken to improve the accuracy of measuring TL such as increasing field diffuseness by using reflecting surfaces, installing near-field array of loudspeakers, and using sound intensity probes on the receiving sides. None of these proved totally satisfactory. This work presents studies on using various approaches to more accurately determine acoustic mitigation of low-frequency transmitted noise. A metal box consisting of a very rigid frame supporting elastic panels on all sides was acoustically excited by a loudspeaker suspended inside of the box. The panels and box interior were designed to have very low-fundamental frequencies. The box was suspended by its frame in an anechoic room. An intensity probe was used to measure the transmitted sound from the box panels. The acoustic performance of various acoustic materials was compared for frequencies below 500 Hz.

10:45

3aNSc3. Acoustic impedance of Earth's ground surface at infrasonic frequencies. Allan J. Zuckerwar (Analytical Services and Mater., Hampton, VA 23666, ajzuckerwar@yahoo.com)

The acoustic impedance of Earth's ground surface is known to have a significant impact upon measurements of outdoor infrasound, e.g., sonic boom rise times. Current acoustic impedance models assume a rigid frame, or a no-slip condition, and thus lead to unbounded values of the real and imaginary parts of the acoustic impedance as the frequency approaches zero. The model of Tarnow [J. Acoust. Soc. Am. **105**, 234–240 (1999)], on the other hand, considers relative motion between a porous frame (soil) and the air in the pores. It is shown here that this model yields finite values of the ground impedance in the limit to zero frequency. Further, extrapolation of selected experimental data to zero frequency by the method of the Richardson extrapolation also leads to finite values of the acoustic ground impedance. Finally, it is shown that the reflection of long-wavelength infrasound from bedrock may reveal resonance effects that will impact the ground impedance.

11:00

3aNSc4. Experience of the measurement and assessment of residential low-frequency noise complaints. David C. Waddington, Andrew T. Moorhouse, and Mags D. Adams (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT,United Kingdom, d.c.waddington@salford.ac.uk)

The practical application of a procedure for the assessment of low-frequency noise (LFN) complaints is described. The development of the assessment method was published recently [J. Acoust. Soc. Am. **126**, 1131–1141 (2009)]. The procedure includes guidance notes and a pro-forma report with step-by-step instructions for field measurements, complemented with an interview-based questionnaire together with an event log completed by the complainant. It does not provide a prescriptive indicator of nuisance but rather gives a systematic procedure to help environmental health practitioners to form their own opinion. Examples of field measurements and application of the procedure are presented. The procedure and examples are likely to be of interest to environmental health practitioners involved in the assessment of LFN complaints. [Work funded by the Department for Environment, Food and Rural Affairs (Defra), UK.]

11:15

3aNSc5. Research into the human response to vibration from railways in residential environments. David C. Waddington, Andrew T. Moorhouse, James Woodcock, Nathan Whittle, Sharron Henning, Eulalia Peris, Gennaro Sica, Andy Steele, Phil A. Brown, and Mags D. Adams (Univ. of Salford, Salford M5 4WT, United Kingdom, d.c.waddington@salford.ac.uk)

This paper describes progress in research being carried out at the University of Salford to develop a method by which human annoyance to vibration in residential environments can be assessed. The objective of this study is to yield a robust relationship between vibration exposure and human response, therefore providing a reliable basis for the development of standards and guidance for the assessment of vibration in residential buildings. The vibration sources to be considered are those affecting residents that are outside their control, such as construction, road, and rail activities. Noise is also a consideration. The protocol involves the measurement of vibration outside and inside individual residences and a social study questionnaire based on face-to-face interviews with householders. Work so far has concentrated on the response of people in their own homes to railway noise and vibration. Approximately 1000 case studies have been obtained, and examples of early field measurements and results are presented. This work is likely to be of interest to acoustical consultants and environmental health officers involved in the assessment of vibration complaints and to planners and practitioners involved in the design of buildings. [Work funded by the Department for Environment, Food and Rural Affairs (Defra) UK.]

11:30

3aNSc6. Measurements of exploding balloon demonstrations. Julia A. Vernon, Kent L. Gee (Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602, julia.vernon@yahoo.com), and Jeffery H. Macedone (Dept. of Chemistry and Biochemistry, Brigham Young Univ., Provo, UT 84602)

Exploding balloons are popular demonstrations in introductory chemistry and physical science classes and as part of outreach programs. However, as impulsive noise sources, these demonstrations constitute a possible hearing damage risk to both the demonstrator and the audience. Measurements of various hydrogen, hydrogen-oxygen, and acetylene-oxygen balloons were made in both a large lecture hall and an anechoic chamber. Condenser microphones (6.35 and 3.2-mm diameters) were placed at various angles, and distances from the balloon and time waveform data were collected at a sampling frequency of 192 kHz. For all balloon sizes tried, hydrogen-only balloons were found to produce peak sound pressure levels less than 140 dB at distances greater than or equal to 2 m. On the other hand, large (but reasonably sized) hydrogen-oxygen and acetylene balloons can result in peak levels exceeding 160 dB at a distance of 2 m, which constitutes a significant hearing risk for unprotected listeners at typical distances.

11:45

3aNSc7. Ultrasonic noise from manmade electical devices. Trevor Jenny (tdwheats@hotmail.com), Brian E. Anderson (Dept. of Phys. and Astron., Acoust. Res. Group, Brigham Young Univ., Provo, UT 84602), and James A. TenCate (Los Alamos Natl. Lab., Los Alamos, NM 87545)

Measurements of ultrasonic noise emitted from manmade electrical devices have been made in an ultrasonic anechoic chamber. Until now, little has been published about the levels and frequency content of ultrasonic noise emitted from electronic devices such as laptops, cell phones, and other electronic devices. The most common source of ultrasonic noise in these devices is dc to dc power supplies. Measurements are made over the range of 20–200 kHz emitted by electronic devices (e.g., laptops). The results of these measurements and the challenges associated with making the measurements will be discussed (including anechoic chamber qualification in the ultrasonic frequency range). [This work has been funded by the Los Alamos National Laboratory.]

Session 3aPA

Physical Acoustics, Structural Acoustics and Vibration, and Noise: Sonic Boom I

Natalia V. Sizov, Cochair
Wyle Laboratories, 241 18th St., Ste. 701, Arlington, VA 22202

Kimberly Lefkowitz, Cochair
CSTI Acustics, 16155 Park Rd., #150, Houston, TX 77084-5100

Victor W. Sparrow, Cochair
Pennsylvania State Univ., Graduate Program in Acoustics, 201 Applied Sciences Bldg., University Park, PA 16802

Invited Papers

8:00

3aPA1. Sonic boom research in NASA's supersonic fundamental aeronautics project. Peter G. Coen (NASA Langley Res. Ctr., MS 264, Hampton, VA 23681)

The Supersonics Project of NASA's Fundamental Aeronautics Program is a broad based effort designed to address the technical challenges associated with flight in the supersonic speed regime. One of the key environmental challenges that must be overcome for successful commercial supersonic flight is sonic boom. The NASA Supersonics Project has chosen this challenge as a focus for its research activity over the next several years. In cooperation with the FAA and the ICAO Committee on Aviation Environmental Protection (CAEP) NASA has initiated the development of a roadmap for research into the community response to sonic booms with noise levels significantly less than those of the Concorde and current supersonic aircraft. The purpose of the roadmap is to define the key research that should be accomplished in order to provide a firm technical basis for future noise-based standards for sonic boom acceptability. This presentation will discuss the elements of this roadmap and outline some of the recent research contributions made by the NASA Supersonics Project.

8:20

3aPA2. Random focusing of nonlinear *N*-waves in fully developed turbulence: Laboratory scale experiment and theoretical analysis. Philippe Blanc-Benon (CNRS, LMFA UMR 5509 Ecole Centrale de Lyon, 69134 Ecully Cedex, France, philippe.blanc-benon@ec-lyon.fr), Mikhail V. Averiyanov (Moscow State Univ., Moscow 119991, Russia), S. Ollivier (CNRS, LMFA UMR, 69134 Ecully Cedex, France), and Vera A. Khokhlova (Phys. Faculty, Moscow State Univ., Moscow 119991, Russia)

The high-amplitude shock wave generated by a supersonic aircraft propagates through the atmosphere toward the ground and generates an acoustic field with non-uniform pressure distributions strongly influenced by atmospheric turbulence. Recent numerical simulations based on generalized KZK-type equation including the effects of moving inhomogeneous media will be discussed. Formation of multiple focusing and defocusing zones is predicted. Nonlinear effects are significant not only in the random focusing zones but also in shadow zones of lower-pressure levels due to scattering of high frequencies from the areas of focusing. A statistical analysis is performed, and the results are compared to experimental data obtained in the controlled laboratory scale experiments conducted in the ECL anechoic wind tunnel. A high-power spark source is used to generate *N*-waves. Correlation length scales and spectra of the turbulent velocity field are measured. Statistical distributions and mean values for peak positive pressure and shock arrival time are obtained and found to be in a good agreement with modeling. In focusing areas, waveforms with amplitudes more than four times higher than those measured without turbulence are observed. Pressure amplitude probability density distributions are shown to possess autosimilarity properties when changing the intensity of turbulence. [Work supported by RFBR, French Government.]

8:40

3aPA3. A comparison of wide-angle and narrow-angle progressive wave equations for modeling sonic boom propagation through turbulence. Andrew Piacsek (Dept. of Phys., Central Washington Univ., 400 E. Univ. Way, Elllensburg, WA 98926-7422, piacsek@cwu.edu), B. Edward McDonald (Naval Res. Lab, Washington, DC 20375), and Victor Sparrow (Penn State Univ., Univeristy Park, PA 16802)

The high-angle formulation of the nonlinear progressive wave equation model (NPE2) [McDonald, Wave Motion **31**, 165–171 (2000)] has been adapted for atmospheric propagation and is used to perform numerical simulations of sonic boom propagation through atmospheric turbulence. Turbulence is incorporated into the model as perturbations of the local ambient sound speed, which have a random spatial distribution determined by a von Karman energy spectrum. Wave forms, peak overpressures, and wavefront profiles computed at fixed ranges from the onset of turbulence are compared to corresponding solutions produced by the original (narrow-angle) version of the NPE in order to quantitatively assess the importance of wide-angle accuracy for sonic boom propagation.

9:00

3aPA4. The influence of non-standard atmospheric conditions on the potential number of acoustic signature occurrences in the continental United States. Joe Salamone (Gulfstream Aerosp., 500 Gulfstream Rd M/S R-07, Savannah, GA 31402)

The potential number of supersonic flight operations and their resulting acoustic signature occurrences are important components for assessing the overall impact of supersonic flight over populated areas. Gulfstream Aerospace has recently published estimates for the number of acoustic signatures that would occur per day, on average, over the continental United States based on actual flight movements from large cabin business jets. The previous analysis of the flight movement data assumed propagation from a supersonic vehicle flying at a constant altitude and Mach number through a standard atmosphere and also assumed great circle routing between the origin and destination airports. The present work will examine the influence that realistic flight profiles and non-standard atmospheric conditions have on the location and extent of the acoustic signature occurrences. The non-standard atmospheric conditions will consist of seasonally averaged, upper-air profiles from nearly 100 locations across the continental United States. The realistic flight profiles will be based on estimates for the climb capability of the supersonic vehicle and possible air traffic routing trajectories.

9:20

3aPA5. Prediction of shaped sonic boom focusing. Anthony R. Pilon (Adv. Development Programs, Lockheed Martin Aeronautics Co., 1011 Lockheed Way, Palmdale, CA 93599, tony.pilon@lmco.com)

The focusing effects of aircraft maneuvers and accelerations on shaped, or minimized, sonic boom signals have been investigated. The spectrally accurate boom prediction methodology of Pilon [AIAA J. **45**, 2149–2156 (2007)] has been enhanced to include the effects of focusing through an implementation of the nonlinear Tricomi equation solver of Auger and Coulouvrat [AIAA J. **40**, 1726–1734 (2002)]. The newly enhanced prediction tool was used to predict boom focus from aircraft accelerations, turns, and dives. Idealized near-field shaped boom signals were used as inputs to the code. Both fully resolved, continuous boom signals and booms containing zero-thickness shock waves were used as inputs to the focusing algorithm, in order to highlight the differences in perceived loudness on the ground. The new tool and results of the study will be used to help design accelerations and maneuvers for future supersonic cruise aircraft.

9:40

3aPA6. Three-dimensional finite-difference time-domain simulation of low-amplitude sonic boom diffraction around a rectangular test structure and building spiking effect. Sang I. Cho and Victor W. Sparrow (Grad. Prog. in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, stc142@psu.edu)

A finite-difference time-domain (FDTD) approach is used for a three-dimensional numerical modeling of pressure loading on the exterior surface of a building due to an incident low-amplitude sonic boom. For these "low-booms," the linearized Euler equations suffice to describe the boom-structure interaction. A second-order centered-difference scheme and fourth-order Runge–Kutta time integration scheme are applied. The FDTD simulation models a field test performed by Virginia Tech in 2008, where a rectangular test structure representing a residential building was instrumented to measure its vibroacoustic response to a boom generated by detonation cord. By matching the geometries of the simulated structure to the actual test structure and using the waveform of a measured boom as the input to the simulation, a direct comparison of numerical and experimental data is made possible. The simulation result agrees well with the experiment and also shows expected pressure doubling and diffraction effects. An interesting observation is made in both numerical and experimental data and is called the "building spiking" effect, which can be attributed to frequency-selective diffraction by the building. [Work supported by NASA. The authors appreciate Virginia Tech and NASA making the 2008 detonation cord test data available for this work.]

10:00—10:20 Break

10:20

3aPA7. A combined ray-tracing and radiosity method for propagation and diffraction of sonic booms in urban landscapes. Kimberly Lefkowitz and Victor W. Sparrow (Pennsylvania State Univ., State College, PA 16801)

Civil supersonic flight over land will only be allowed when sound levels are low enough to be acceptable to the public, and thus the sound field inside buildings needs to be well understood. To understand the field inside buildings, the sound field outside the buildings as well as the diffraction and loading on the building must be determined. A model to propagate sonic booms in urban landscapes was developed using a combined ray tracing and radiosity method. This model includes both diffuse and specular reflections for a single building configuration. The impact of diffuse reflections on the propagating and diffracting sound field will be examined, and multiple building configurations are also considered. The model results will be compared with data from NASA's SonicBOBS (sonic booms on big structures) experiment conducted in September of 2009. [Work supported by NASA.]

10:40

3aPA8. Vibro-acoustic response of a simplified residential structure exposed to sonic booms. Part I: Numerical model. Marcel Remillieux, Joseph Corcoran, Ryan Haac, Ricardo Burdisso (Virginia Tech, 142 Durham Hall, Blacksburg, VA, 24061, mremilli@vt.edu), and Georg Reichard (Myers-Lawson School of Construction, Virginia Tech)

This paper presents a numerical model to predict the vibro-acoustic responses of simplified residential structures exposed to sonic booms. The model is validated experimentally in a companion paper. The dynamics of the fluid-structure system, including their interaction, is computed in the time domain using a modal-decomposition approach. In the dynamic equations of the system, the structural displacement is expressed in terms of summations over the *in vacuo* modes of vibration. The pressures inside the interior volumes are expressed as summations over the acoustic modes of rooms with perfectly reflecting surfaces. The structural modes are computed nu-

merically using the finite element method. A shell element was specifically derived to model the structural components of typical residential buildings, e.g., plaster-wood walls, windows, and doors. The acoustic modes are computed for rectangular geometries using analytical expressions. Using modal decomposition, the dynamics of the fluid-structure system may be formulated by a finite set of ordinary differential equations (modal equations). These equations are then integrated with a Newmark algorithm to solve for the vibro-acoustic response of the system in the time domain. The system response may also be predicted in the frequency domain, by taking the Fourier transform of the time-domain response.

11:00

3aPA9. Vibro-acoustic response of a simplified residential structure exposed to sonic booms. Part II: Experimental validation. Marcel Remillieux, Joseph Corcoran, Ryan Haac, RicardoBurdisso Burdisso (Virginia Tech, 142 Durham Hall, Blacksburg, VA 24061, mremilli@vt.edu), and Georg Reichard (Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA 24061)

In a companion paper, a model was derived to predict the vibro-acoustic responses of simplified residential structures exposed to sonic booms. In the present paper, the experimental validation of the numerical model is presented. First, the experimental setup is described, including the structure, instrumentation, and external pressure loading. The structure was a single room made of plaster-wood walls and includes two double-panel windows and a door. The structure was extensively instrumented with accelerometers and microphones to record its vibro-acoustic response. Sonic booms with realistic amplitudes and durations were generated by an explosive technique. Subsequently, numerical predictions on the vibration of the structure and pressures inside the room are compared to experimental data, showing a fairly good agreement overall.

11:20

3aPA10. Distributed field measurements of low amplitude sonic booms. Joseph Gavin (Gulfstream Aerosp., 32 Lummus Dr, Savannah, GA 31408, joseph.gavin@gulfstream.com), Jack Arnold, Bryan Nadeau (Natl. Instruments, Austin, TX), Jacob Klos (NASA Langley Res. Ctr., Hampton, VA. 23681), and Edward A. Haering, Jr. (NASA Dryden Flight Reearch Ctr., Edwards, CA)

Quiet supersonic flight over land appears increasingly viable, based on significant advancements in aircraft shape optimization. One of the next critical milestones will involve community scale surveys, measuring the public response to events that sound as benign as distant thunder. The present paper describes a measurement infrastructure to reduce the field costs associated with those tests. Specifically, a scalable grid of NI CompactRIO data acquisition systems has been developed and deployed in a risk reduction test. Highlights of the system include communications via a wireless rf network and synchronization using low cost GPS receivers. This paper highlights some of the development challenges, solutions, and recent results from component and system level risk reduction testing at the NASA Dryden Flight Research Center.

11:40

3aPA11. Detailed finite element analysis of sonic boom transmission into residential building structures. Beom Soo Kim , W. Victor Sparrow (Graduate Program in Acoust., 201 Appl. Sci. Bldg., University Park, PA 16802, buk104@psu.edu), V. Natalia Sizov, and T. Ben Manning (Wyle Labs., Arlington, VA 22202)

Finite element modeling is a powerful and cost-effective tool for solving problems in structural dynamics and acoustics, particularly at low frequencies. In the present work the transient responses of the sonic boom transmission from outdoors to indoors were studied. Detailed finite element models of individual rooms from two residential structures were built. The simulated geometries were matched to the physical geometries of two houses exposed to sonic booms that were heavily instrumented by NASA in the summers of 2006 and 2007. The residential building walls were composed of studs and sheathing, and these were modeled using MSC NASTRAN/PATRAN software. To match the normal mode frequencies of the walls, it was necessary to not only match material coefficients but also the actual connections between the studs and the sheathing. The fluid structure coupling between the walls and the air was utilized to analyze the acoustic response of the sonic boom in the room interior. In order to verify the correctness of the model, the vibration responses of the structure and the acoustic cavity were compared with measured data. [Work supported by NASA.]

Session 3aPP

Psychological and Physiological Acoustics: Sound Source Localization

Joseph W. Hall, III, Chair
Univ. of North Carolina, Dept. of Otolaryngology, CB 7070, Chapel Hill, NC 27599

Contributed Papers

9:00

3aPP1. Sound source localization of a dipole by the plainfin midshipman fish (*Porichthys notatus*). David G. Zeddies (Marine Acoust. Inc., 41 Fairfax Dr., Arlington, VA 22203), Richard R. Fay (Loyola Univ.-Chicago, Chicago, IL 60626), Peter W. Alderks, Andrew Acob, and Joseph A. Sisneros (Univ. of Washington, Seattle, WA 98195)

Localization of a dipole sound source was studied in female plainfin midshipman fish (*Porichthys notatus*). Experiments were conducted and videotaped in a 3.65-m-diameter tank using a dipole underwater speaker system placed near the center of the tank. The sound was a 90-Hz tone, approximately the fundamental frequency of the male's advertisement call. Pressure and particle motion components of the sound field were mapped with 9-cm resolution. Pressure was measured using an eight-hydrophone array, and particle motion vectors calculated from the pressure gradients. Mapping confirmed that the projector was operating as a dipole. Gravid fish were released 70 cm from the sound source at two different positions relative to the dipole axis: one near the dipole axis and one near the pressure null axis. Twenty-five positive responses were recorded from each release site. The phonotactic response pathways along the dipole axis consisted of slightly curved tracks to the sound source, whereas pathways from the null axis consisted of greatly curved tracks to the source that followed the particle motion vectors. Results confirm that fish can locate a dipole sound source and are sensitive to the direction of acoustic particle motion. [Work supported by NSF.]

9:15

3aPP2. Generating correlated noise. William M. Hartmann and Yun Jin Cho (Dept. of Phys. and Astronomy, Michigan State Univ., East Lansing, MI 48824)

Research on binaural hearing frequently presents partially correlated noise to a listener's two ears. Experimentally, there are two classic methods to generate noise with a controlled amount of cross-correlation: the *two-generator* method and the *three-generator* method [J. C. R. Licklider and E. Dzendolet, Science, **107**, 121–124 (1948)]. Analytic formulas were developed and numerical experiments were done to study the properties of the two methods, particularly to determine how well the actual correlations approach targeted correlations. It was found that both methods give very similar results, with variance that is greatest when the targeted cross-correlation is zero and which decreases monotonically to zero as the targeted value increases to 1.0. The variance decreases inversely as the first power of the number of degrees of freedom in the noise, whatever the targeted correlation. Paradoxically, a far less popular method, the *symmetrical generator* method, is far more successful. The variance is zero when the targeted cross-correlation is either zero or 1.0, and the variance decreases inversely as the square of the number of degrees of freedom. The symmetrical generator method is a reasonable alternative to the summing of digitally orthogonalized noises. [Work supported by the NIDCD DC-00181.]

9:30

3aPP3. A binaural localization model that resolves front-back confusions through head movements. Jonas Braasch, John T. Strong, and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, braasj@rpi.edu)

It is well known that head movements are instrumental in resolving front/back confusions in human sound localization. A mechanism for a binaural model is proposed here to extent current cross-correlation models to compensate for head movements. The algorithm tracks sound sources in the head-related coordinate system (HRCS) as well as in the room-related coordinate system (RRCS). It is also aware of the current head position within the room. The sounds are positioned in space using an HRTF catalog at 1 deg azimuthal resolution. The position of the sound source is determined through the inter-aural cross-correlation (IACC) functions across several auditory bands, which are mapped to functions of azimuth and superposed. The maxima of the cross-correlation functions determine the position of the sound source, but unfortunately, usually two peaks occur—one at or near the correct location and the second one at the front/back reversed position. When the model is programed to virtually turn its head, the degree-based cross-correlation functions are shifted with current head angle to match the RRCS. During this procedure, the IACC peak for the correct hemisphere will prevail if integrated over time for the duration of the head movement, whereas the front/back reversed peak will average out.

9:45

3aPP4. Modeling the precedence effect for multiple echoes. Matthew J. Goupell and Ruth Y. Litovsky (Waisman Ctr., Univ. of Wisconsin-Madison, Madison, WI 53705)

A typical and controlled way to study the precedence effect has been to use two pairs of binaural click stimuli with different lead-lag delays. For lead-lag delays between 1 and 5 ms, interaural differences in the lead click pair are more easily discriminated than those in the lag click pair. This work attempts to move from the classic, but unrealistic, two-source paradigm to a more realistic situation, much like real rooms in which multiple echoes occur. Using a three-source paradigm, listeners were presented three click pairs (lead, lag1, and lag2) via headphones. Discrimination of directional changes in each of the three click pairs was measured with a 2AFC task for delays between 0 and 130 ms. Results show that due to interactions amongst the three sources "echo thresholds" were affected in a non-monotonic way. Temporal order effects were observed for long delays, which led to poorer lead discrimination, and better lag2 discrimination. Summing localization was observed for short delays between lead and lag1, and lag1 and lag2. A simple weighting model [Shinn-Cunningham *et al.*, J. Acoust. Soc. Am. **93**, 2923–2932 (1993)] was adapted to include this multiple source data and the complex relationships between the three sources. [Work supported by NIH-NIDCD R01 DC003083—R.Y. L.]

10:00

3aPP5. Investigation into and modelling of head movement for objective evaluation of the spatial impression of audio. Chungeun Kim, Russell Mason, and Tim Brookes (Inst. of Sound Recording, Univ. of Surrey, Guildford GU2 7XH, United Kingdom)

Research was undertaken to determine the nature of head movements made when judging spatial impression and to incorporate these into a system for measuring, in a perceptually relevant manner, the acoustic parameters which contribute to spatial impression: interaural time and level differences and interaural cross-correlation coefficient. First, a subjective test was conducted that showed that (i) the amount of head movement was larger when evaluating source width and envelopment than when judging localization

and timbre and (ii) the pattern of head movement resulted in ear positions that formed a sloped area. These findings led to the design of a binaural signal capture technique using a sphere with multiple microphones, mounted on a simulated torso. Evaluation of this technique revealed that it would be appropriate for the prediction of perceived spatial attributes including both source direction and aspects of spatial impression. Reliable derivation of these attributes across the range of ear positions determined from the earlier subjective test was shown to be possible with a limited number of microphones through an appropriate interpolation and calculation technique. A prototype capture system was suggested as a result, using a sphere with torso, with 21 omnidirectional microphones on each side. [Work supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant No. EP/D049253.]

10:15—10:30 Break

10:30

3aPP6. Tuning of head-related transfer functions using principal component analysis. Kimberly J. Fink and Laura R. Ray (Thayer School of Eng., Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755)

Modeling head related impulse responses (HRIRs) to determine head-related transfer functions (HRTFs) can be used to create a virtual auditory display through head phones. However, non-individualized HRTFs can create errors in sound localization and perception of the source as coming from inside vs outside of the head. We present a method, based on model-order reduction and principal component analysis, to "tune" generalized HRTFs derived from the CIPIC database of HRIRs in order to individualize an HRTF. The impulse response for each individual in the CIPIC database is modeled at given azimuth locations on the horizontal plane. Model order reduction is used to eliminate poles and zeros of the identified HRTF at high frequencies, while preserving the frequency response below ~8000 Hz. Principal component analysis is then performed on the reduced-order models in order to reduce the dimensionality of the dataset and identify a subset of weighting variables that are adjusted during HRTF customization. The sensitivity of weights on the frequency response is also studied in order to correlate weights with HRTF features. A prototype auditory display has been developed that allows a user to tune HRTFs in real-time, while listening to a 44.1-kHz source presented from a chosen azimuth.

10:45

3aPP7. Effects of individualized headphone equalization on front/back discrimination of virtual sound sources. Abhishek Guru, William L. Martens, and Doheon Lee (Faculty of Architecture, Design and Planning, Univ. of Sydney, New South Wales 2006, Australia, agur4608@uni.sydney.edu.au)

Individualized head related transfer functions (HRTFs) were used to process brief noise bursts for a two-interval forced choice front/back shift between virtual sound source locations presented via two models of headphones, the frequency responses of which could be made nearly flat for each of 21 listeners using their own individualized headphone equalization filters. In order to remove tone coloration differences between virtual sources processed using individualized HRTFs measured in front or in back of each listener, the spectral centroid of the "back" source was adjusted to more nearly match that of the source processed using the "front" HRTF. This manipulation resulted in chance levels of discrimination performance for 12 out of 21 listeners. For the remaining nine listeners showing good discrimination, the virtual sources presented using individualized headphone equalization supported significantly better front/back discrimination rates than did virtual sources presented without correction to headphone responses.

11:00

3aPP8. Reverberation disrupts spatial selective auditory attention. Dorea R. Ruggles and Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215)

When there are multiple, competing speech streams, listeners can selectively attend to a desired target using spatial cues like interaural time differences (ITDs). However, reverberation smears fine temporal information needed to encode ITDs. Reverberation may therefore interfere with the ability to selectively attend to a target from a particular direction. Subjects were asked to report a stream of digits simulated to be directly in front of them in the presence of two masker streams symmetrically positioned at 15 deg left and right. Because the maskers were statistically identical to the target (except in direction), listeners had to focus spatial attention on the target while ignoring the masker streams. Three levels of reverberation were simulated using a rectangular room model to measure the influence of reverberation on performance. Initial results showed that spatial selective attention degraded as reverberation increased. However, overall performance varied dramatically from listener to listener in ways that were not predicted by age or memory span. The current study explores whether individual differences are predicted by performance on two discrimination tasks that rely on signal fine temporal structure: frequency modulation (a monaural task) and ITD (a binaural task). [NIDCD, ONR, and NSF provided funding to support this work.]

11:15

3aPP9. The effects of intervening interference on working memory for sound location. Aurora Grossmann, Dennis Ries, and Traci Hamilton (School of Hearing, Speech, & Lang. Sci., Ohio Univ., Grover Ctr., Athens, OH 45701, ag303003@ohio.edu)

Inter-aural phase differences were used to study working memory for perceived auditory position quantified as a change in the difference limen (DL) in equivalent auditory angle across 300-, 5000-, and 15 000-ms retention intervals. Data were obtained for the medium and long intervals both in the presence and absence of intervening tones. Intervening stimuli within the medium and long inter-comparison intervals produced a significant increase in the DLs compared those obtained in the corresponding quiet conditions. The DL for the 300-ms interval was roughly equivalent to that obtained for the medium interval without intervening tones while that obtained for 15 000-ms interval was significantly greater than that obtained for either of the shorter intervals. The results suggest that the temporal decay of information within AWM of a listerner regarding the location of a sound within their environment is so gradual that it can be maintained in trace memory for tens of seconds in the absence of intervening acoustic signals. Conversely, the presence of interpolated sounds within the retention interval may facilitate the use of context memory, resulting in a less detailed, but relevant representation of the location that is resistant to further degradation.

Session 3aSA

Structural Acoustics and Vibration and Physical Acoustics: Damping Mechanisms

J. Gregory McDaniel, Chair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Chair's Introduction—9:10

Invited Papers

9:15

3aSA1. Damping in auxetic materials and structures. Fabrizio Scarpa (Dept. of Aerosp. Eng., Univ. of Bristol, 83 Woodland Rd., Bristol BS8 1US, United Kingdom, f.scarpa@bris.ac.uk)

Negative Poisson's ratio (NPR) (otherwise called auxetics) materials and structures feature enhanced energy absorption under cyclic loading both at low- and high-frequency excitations. We show in this paper experimental and simulation results related to open cell PU-PE foams, honeycombs, and composites, all exhibiting a negative Poisson's ratio value. The open cell auxetic foams provide a significant damping capacity at low frequency and high preload as well as under broadband excitation compared to the conventional foams from which they are derived. We show also that the acoustic absorption properties of the auxetic foams are significantly enhanced compared to conventional PU-PE under 500 Hz, and their loss factor increased using their specific shape memory property. We describe also the vibration damping properties of NPR honeycombs with different unit cell topologies and the damping capacity of NPR carbon fiber composites under flexural cyclic loading.

9:35

3aSA2. Intrinsic damping, relaxation processes, and internal friction in vibrating systems. Allan D. Pierce (Dept. of Mech. Engr., Boston Univ., Boston, MA 02215, adp@bu.edu)

Vibrations lose energy because of (1) coupling of the system to other systems or to the environment or of (2) intrinsic damping within the system itself. The present paper attempts to survey the current state of understanding for the principal mechanisms of the latter type with explicit physics (rather than purely phenomenological considerations), as simply as practical. Some mechanisms are evidenced in the attenuation of waves in unbounded homogeneous media, others are caused by the presence of boundaries. Some explicitly require quantum-mechanical concepts for their explanation, others can be explained in terms of classical mechanics and thermodynamics. Specific mechanisms that are to be discussed include those associated with thermal conductivity, grain and domain boundary effects, interstitial atom diffusion, relaxation by diffusion, inter-grain diffusion, micro-eddy current losses, hysteresis losses, thermoelastic internal friction, thermal damping associated with transverse vibrations of thin bodies, phonon-phonon interactions, phonon-electron interactions, internal fluids, and dislocations. The concept of relaxation often provides a convenient unified description. Rather than use the local averaged particle displacement as a sole description of the local vibrations, one considers additional local variables, loosely referred to as hidden variables. Each such variable, when disturbed, relaxes to its equilibrium value with a characteristic relaxation time.

9:55

3aSA3. Optimization of local damping treatments for plate vibrations. J. Gregory McDaniel (Dept. Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

This work considers passive damping treatments that dynamically interact with plates over areas that are much smaller than a flexural wavelength. Each treatment is modeled as a frequency-dependent admittance presented to the plate at a point. The frequency-dependence is required to satisfy the causality and minimum phase conditions, which are implicitly enforced via a specialized Fourier series in frequency. The Fourier coefficients, as well as the positions of the attachments, are regarded as optimization parameters to be varied so as to globally decrease plate vibrations over a frequency band of interest. Various cost functions are considered, such as the sum of modal loss factors over the band as well as acoustic radiation efficiencies. Examples involving a simply supported plate excited by a point force illustrate the approach. [Work supported by ONR under Grant No. N000140810531.]

10:15—10:30 Break

10:30

3aSA4. Radiation damping and scattering reduction by transformation acoustics. Andrew Norris (Dept. of Mech. and Aerosp. Engng., Rutgers Univ., NJ 08854, norris@rutgers.edu)

The idea in transformation acoustics is to map one region into another so that the new region has the same acoustic appearance to observers as the original. It is analogous to transformation optics and is closely related to devices for cloaking, but it is not restricted to cloaking. This talk will explore the implications of transformation acoustics for radiation damping and scattering. Mechanisms for realizing transformed regions, which we call acoustic metafluids, will be described. The objective in these materials is to bend the waves so that energy can be steered around a given region, and thereby reduce the scattering strength. But it is not as obvious how these materials influence the radiation from an active source inside. In this talk we will discuss the implications of transformation acoustics on radiation using fundamental concepts, including acoustical reciprocity, mechanical impedance, and energy flux.

Contributed Papers

10:50

3aSA5. Radiation reduction using layers of three fluids. Adam J. Nagy and Andrew N. Norris (Dept. of Mech. and Aerosp. Engng., Rutgers Univ., Piscataway, NJ 08854, adnagy86@gmail.com)

Significant reduction in target strength and radiation signature can be achieved by surrounding an object with multiple concentric layers comprised of three acoustic fluids. The idea is to make a finely layered shell with the thickness of each layer defined by a unique transformation rule. The shell has the effect of steering incident acoustic energy around the structure and, conversely, reducing the radiation strength. The overall effectiveness and the precise form of the layering depend upon the densities and compressibilities of the three fluids. The best results are obtained if one fluid has density equal to the background fluid, while the other two densities are much greater and much less than the background values. Optimal choices for the compressibilities are also found. Simulations in two dimensions and three dimensions illustrate the effectiveness of the three-fluid shell.

11:05

3aSA6. Transmission loss of membrane-type locally resonant acoustic materials. Christina Naify (Dept. of Mater. Sci., Univ. of Southern California, 3651 Watt Way, VHE 602, Los Angeles, CA 90089), Chia-Ming Chang, Bill Carter, Geoff McKnight (HRL Labs., Malibu, CA 90265-4797), and Steve Nutt (Univ. of Southern California, Los Angeles, CA 90089)

Membrane-type acoustic metamaterials were fabricated, characterized, and analyzed. Thin plates which obey the acoustic mass law have low transmission loss at low frequencies. Acoustic metamaterials with negative dynamic mass density have been shown to demonstrate a significant (ten times) increase in transmission loss (TL) over mass law predictions for a narrow band (100 Hz) at low frequencies (100–1000 Hz). The peak TL frequency can be tuned to specific values by varying the membrane and mass properties. In this work, transmission loss magnitude as a function of frequency was measured for variations in the mass magnitude, mass placement on the membrane, membrane size, and membrane thickness using an impedance tube setup. The dynamic properties of membranes constructed from different materials and thicknesses were measured and compared to the results of coupled field acoustic-structural FEA modeling to understand the role of tension and element quality factor. In addition, the behavior and limitations of stacking multiple membranes in series to increase the reflection bandwidth were explored.

11:20

3aSA7. Damping mechanism concepts in ocean wave energy conversion: A simplified model of the Pelamis converter. Amadou G. Thiam and Allan D. Pierce (Boston Univ., Boston, MA 02215, thiam@bu.edu)

The conversion of mechanical energy to another energy form is often built into a structural acoustics model by adding some damping mechanism to the model. Also, in ocean wave energy conversion, one damping mechanism, which is occasionally of great importance in the determination of the limiting upper efficiency of a device, is the re-radiation of wave energy back into the ocean. Present paper focuses on a simplified model of the Pelamis wave energy converter, with the model consisting of a modified Euler–Bernoulli beam oriented head-on to incoming ocean waves and with energy conversion accounted for by a damping term, where the additional bending moment is proportional to the time derivative of the beam curvature. The wave energy converted is ideally equal to the energy dissipated by this term. The beam is taken as having circular cross-section with a mass density that is one-half that of water. An incident wave causes a transverse force along the bottom of the beam so that it heaves up and down in a snake-like fashion, and this heaving causes ocean waves to be re-radiated back into the ocean, causing an additional apparent damping. Consideration is limited to a linear analysis and to a deep-water ocean.

11:35

3aSA8. Rapid step response with limited ringing and light damping. Joseph Vignola and John Judge (Dept. of Mech. Eng., The Catholic Univ. of America, 620 Michigan Ave., NE Washington, DC 20064, vignola@cua.edu)

The time domain step response of an underdamped simple harmonic oscillator can be characterized as having two parts. The first is a transition from the initial position where the structure is accelerated toward a target position by a step force that introduces energy to the system. The second is a ringdown process, involving decaying oscillations about the target position as energy is removed from the structure. This study reports how an array of substantially smaller subordinate oscillators attached to the primary structure can dramatically reduce the time constant of the ringdown (the settling time) without slowing the first phase (rise time) in the way added classical damping does. In addition to having a more rapid initial response, a lightly damped oscillator with such an array of attachments requires less energy to change states than a more heavily damped structure. The approach is based on the concept of apparent damping, where vibration energy is transferred into and trapped by the set of attachments. Numerical results show that the distribution of the individual masses, isolated natural frequencies, and quality factors Q of the attachments, referred to as a subordinate oscillator array, can be tailored to achieve the rapid step response.

11:50

3aSA9. Application of time windowing to spatial maps of damping. Hande Öztürk and J. Gregory McDaniel (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Some of the most challenging vibration control problems involve the selection and placement of frequency-dependent damping devices, such as multilayered viscoelastic materials. In complex structures involving the simultaneous use of several passive damping devices, it is difficult to determine the effectiveness of each device directly from measurements. This question motivates the present analysis, which seeks to quantify the power being dissipated by the structure at selected locations for a given force distribution. The analysis begins with an expression for power flow to the structure from a specified force distribution. This expression involves the real part of the admittance matrix looking into the structure at a set of measurement points. By expanding this matrix as a sum over damping elements that mechanically connect measurement points, one can spatially map the power flow into the structure. While thermodynamics requires that this power flow to the structure be equal to the power dissipated by the structure, it does not require equality of the spatial distributions. In order to understand the spatial distribution of power dissipation, time windowing of the impulse responses is introduced. Examples are developed to explore the effects of the time window. [Work supported by ONR under Grant No. N000140810531.]

Session 3aSC

Speech Communication: Exploring the Relationship Between Cognitive Processes and Speech Perception

Amee Shah, Chair
Cleveland State Univ., Speech and Hearing, 2121 Euclid Ave., Cincinnati, OH 44115

Chair's Introduction—8:00

Invited Papers

8:05

3aSC1. Perceptual expectations, attention, and speech recognition. Howard C. Nusbaum (Dept. of Psych., The Univ. of Chicago, 5848 S. Univ. Chicago, IL 60637, hcnusbaum@uchicago.edu)

Research has demonstrated substantial plasticity in the adult speech perceiver. When listeners receive feedback about the classification of a speech signal, this feedback can shape attention affecting phoneme perception. This demonstrates that it is possible to shape perceptual attention prospectively changing subsequent speech perception. These changes can also reduce the cognitive load of speech perception. The effect of expectations on speech perception has been reported for a wide range of studies from sinewave speech perception to phonetic categorization and effects of talker variability. Neural changes occurring during learning and the neural effects of expectations both appear to involve regions of motor cortex implicated in speech production. These neural changes suggest a model of speech perception similar to analysis-by-synthesis in which sensory processes are tuned by activity within the speech motor system. Although such sensory-motor interactions have been implicated in speech perception due to coarticulation and the encoding of speech into sound, the role of expectations in speech perception may extend beyond segmental perception. We consider evidence that the interpretation of prosodic signals also depends on perceptual expectations, suggesting that the role of expectations may be to constrain recognition given the many-to-many relationship that holds broadly in speech.

8:35

3aSC2. What do comparisons between younger and older adult listeners tell us about speech processing? Margaret K. Pichora-Fuller (Dept. of Psych., Univ. of Toronto, 3359 Mississauga Rd. N, Mississauga, ON L5L 1C6, Canada, k.pichora.fuller@utoronto.ca)

The perception of speech, the recognition of words, and the understanding of spoken language involve the dynamic and interactive processing of cues provided by the incoming signal and information stored in memory. Even when accuracy is high, the relative contributions of bottom-up and top-down processes may explain variations in the speed and effort required when listening to speech. Listening is fast when the quality of the incoming signal is optimal, but it is slowed as signal quality is reduced. Likewise, listening can be speeded when expectations constrain the likely alternatives or when priming implicitly facilitates the recognition of the signal, whereas it can be slowed if the context is incongruent with the signal or if context is used to resolve ambiguities or repair misperceptions in a compensatory fashion. Within-subjects comparisons on off-line and on-line measures in different listening conditions, including simulations of auditory aging and hearing loss, are used to investigate how listening effort varies and how listening is speeded or slowed depending on signal-driven and knowledge-driven factors. Comparisons between younger and older participants are used to evaluate how long-standing reductions in auditory temporal processing and compensatory changes in brain organization may alter how signal-driven and knowledge-driven processes interact.

9:05

3aSC3. The talker-specific nature of spoken language processing. Lynne C. Nygaard (Dept. of Psych., Emory Univ., 36 Eagle Row, Atlanta, GA 30322, lnygaar@emory.edu)

During spoken communication, listeners must contend with a physical speech signal that carries information not only about the linguistic content of a talker's utterance but also about a host of talker-specific attributes such as talker identity and emotion. Traditionally, the recovery of linguistic structure has been thought to be independent of the recovery of these indexical or surface characteristics of speech. Recent research, however, suggests that listeners remember the perceptual quality of spoken words, flexibly adapt to informative indexical variation in speech, and use this variation to guide their understanding of spoken words. This talk will review evidence for indexical specificity effects across a range of perceptual and memory tasks and will examine the linguistic, attentional, and cognitive constraints that may mediate the role of surface form in spoken language processing. Implications of these effects for accounts of linguistic representation and processing will be discussed. [Work supported by NIDCD.]

9:35

3aSC4. Interpreting speech in context. Mark Pitt (Dept. of Psych., Ohio State Univ., 1835 Neil, Ave., Columbus, OH, 43210, pitt.2@osu.edu)

One of the challenges in understanding how humans recognize spoken language is to reconcile the robustness of verbal communication with the many forms of ambiguity and variability in the speech signal. How much of the discrepancy is due to our incomplete understanding of the acoustics of speech (i.e., what is the critical information) versus the recruitment of mental processes (e.g., memory) to aid recognition? My approach to these explanations has been to study the processing of words that have been degraded in controlled ways (e.g., digital manipulation) and by using conversation-style speech that contains naturally produced ambiguities. Recognition is measured in isolation and in sentences using a variety of tasks whose purpose is to measure the source, magnitude, and time course of ambiguity resolution. I will present an overview of this work in the context of a few experimental settings, highlighting methodological issues along the way. Together, the findings suggest that word recognition is a highly fluid process that depends on the rapid integration of multiple sources of information.

10:05—10:20 Break

10:20

3aSC5. Investigating whether ease of lexical processing affects listeners' ratings of the strength of talkers' foreign accent. Amee Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 429, Cleveland, OH 44115) and Conor McLennan (Dept. of Psych., Cleveland State Univ., 2121 Euclid Ave., CB 175, Cleveland, OH 44115)

Many research studies in spoken word recognition have examined the role that different types of surface variability have on listeners' perception of spoken words (e.g., influence of same versus different talker manipulations). In our ongoing research program, we are investigating the other direction of this relationship. Our initial two experiments investigated whether linguistic complexity affects listeners' overt subjective impressions of the strength of talkers' foreign accents. Ease of lexical processing was manipulated in two ways, namely, sentence context and long-term repetition priming. In the sentence context experiment, participants made accent ratings on the last word of each sentence presented to them; this word was either related or unrelated to the sentence context. In the long-term repetition priming experiment, participants made their accent ratings to isolated spoken words, which they either had (primed) or had not (unprimed) heard earlier in the experiment. Stronger accent ratings in the unrelated sentence context and unprimed conditions were predicted. Response time to provide their accent rating responses to sentence context and priming manipulations was also studied. Our results should lead to a greater understanding of the role that lexical processing plays in listeners' perception of foreign accents and perhaps other types of surface variability.

10:50

3aSC6. Interactive processing in speech perception: The facts of feedback. Daniel Mirman (Moss Rehabilitation Res. Inst., 1200 W. Tabor Rd., Philadelphia, PA, 19141)

Context plays a critical role in human speech processing. Rapid, efficient integration of context and input is accomplished by interactive processing: bottom-up input information and top-down context information work together to constrain the percept. One domain where this is most clear is the effects of word context on perception of speech sounds. Behavioral experiments and computational model simulations show that interactive processing confers both benefits (robust perception in noisy environments and tuning in response to changes in input patterns) and costs (errors and delays in perception when the input is inconsistent with the context). Context effects are also prevalent is the constraining effect of global communication context on activation of different meanings of ambiguous words (i.e., homophones): when only highly imageable meanings are consistent with the context (a word-to-picture matching task), concrete noun meanings (e.g., the tree-related meaning of bark) become more activate than less imageable meanings (e.g., the dog-related meaning of bark). These findings are consistent with an interactive graded constraint satisfaction view of speech perception in which bottom-up input and top-down context simultaneously constrain the final percept.

11:20—12:00 Panel Discussion

Session 3pAA

Architectural Acoustics and Noise: Primary and Secondary School Special Function Spaces II

Robert C. Coffeen, Chair
Univ. of Kansas, School of Architecture and Urban Design, Lawrence, KS 66045

Invited Papers

1:00

3pAA1. Acoustical design of special purpose rooms in schools: Cafeterias and cafetoriums. Gary W. Siebein (Univ. of Florida School of Architecture, P.O. Box 115702, Gainesville, FL 32611), Hyun Paek, Chris P. Jones, and Reece Skelton (Siebein Assoc., Inc., Gainesville, FL 32611)

Design strategies for cafeterias and cafetoriums in elementary, middle, and high schools based on an impulse response-based theory are presented through a series of case studies of existing, renovated, and designed rooms. The strategies include use of sound absorbing materials to control noise build-up for dining areas; integrated design of stage, canopy, shells (where used), and sound reinforcement systems to project sounds to the audience in a multi-use situation; controlling room volume and absorption to allow early sound reflections for clear hearing; providing for clear verbal communication between teachers and students with the audience; and limiting background and intruding noise levels. The impulse response-based measures are reduced to a series of architectural systems that can be implemented using alternate construction systems to meet budget requirements. Standard design concept sketches communicate the intent of the acoustical design to architects, school board members, design builders, and music faculty.

1:20

3pAA2. A new sound system in a reverberant high school gymnasium. Shane J. Kanter, Stephanie L. Hoeman, and Robert C. Coffeen (Dept. of Architecture, School of Architecture, Design and Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045, shanekanter@gmail.com)

A new and well designed sound system was recently installed in a very reverberant high school gymnasium. While the new sound system performs much better than the previous system, would its performance be significantly better if the gymnasium reverberation time and undesirable sound reflections were properly controlled? This situation is investigated with binaural audio recordings of the new sound system with the existing room acoustics and by computer modeling and computer auralization with existing acoustical conditions and with improved room acoustics.

1:40

3pAA3. Design constraints and opportunites of two similar-function auditoria. Robert Lee and Andrew Miller (BAi, LLC, 4006 Speedway, Austin, TX 78751, robert@baiaustin.com)

Two high school auditoria are examined, one new construction and one renovation. The multi-purpose uses, size, and scope of both rooms are very similar, but they demanded very different design approaches due to the varying constraints and opportunities presented by the site, the owner, the existing building, and other factors.

2:00

3pAA4. Case studies: Auditorium, gymnasium, and gymnatorium. David S. Woolworth (Oxford Acoust., Inc., 356 CR 102, Oxford, MS 38655, dave@oxfordacoustics.com)

Three case studies of multipurpose rooms in schools with restricted budget. Areas covered include considerations with regard to function, re-use of materials, acoustic treatments fit under the school board's scope, hearing impaired provisions, noise issues, and electroacoustic design.

Session 3pAB

Animal Bioacoustics: General Topics in Animal Bioacoustics (Poster Session)

Peter Marvit, Cochair
Univ. of Maryland, Dept. of Psychology, College Park, MD 20742

Susan E. Parks, Cochair
Pennslyvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Contributed Papers

All posters will be on display from 1:00 p.m. to 3:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:00 p.m. and contributors of even-numbered papers will be at their posters from 2:00 p.m. to 3:00 p.m.

3pAB1. Acoustic changes in the vocal signals of grasshopper sparrows, *Ammodramus savannarum*, during early development. Bernard Lohr and Lacey Laudick (Dept. of Biological Sci., Univ. of Maryland Baltimore County, Baltimore, MD 21250)

Grasshopper sparrows produce unusually high-pitched vocalizations for birds, with fundamental frequencies of species typical songs in the range of 7–9 kHz. The characteristics of vocalizations produced early in life, however, are relatively understudied in this and related species. In this study, the development of vocal signal structure and vocal behavior in grasshopper sparrows was investigated in the field immediately after birds hatched. Acoustic characteristics of nestling begging calls were measured throughout the course of the 8–9-day nestling period. Starting at day 6 post-hatch, nestling calls changed qualitatively, incorporating frequency-modulated elements that resembled some parts of the adult vocalizations. With increased nestling age, there was also an increase in the average duration, average acoustic frequency (pitch), and average amplitude of nestling calls. Overall, our results showed that the complexity of nestling vocal signals increased during early development, with vocalizations exhibiting some adult-like properties by the end of the nestling period.

3pAB2. Echolocation of moving objects by using the frequency modulated sound. Ikuo Matsuo (Dept. of Information Sci., Tohoku Gakuin Univ., 2-1-1 Tenjinzawa, Izumi-ku, Sendai, Miyagi 981-3193, Japan and Bio-Navigation Ctr., Doshisha Univ., 1-3 Miyakodani, Tatara, Kyotanabe, Kyoto 610-0321, Japan, matsuo@cs.tohoku-gakuin.ac.jp)

Using the echolocation, bats can capture moving objects in real three-dimensional space. Bats emit the frequency modulation sound and can identify objects with an accuracy of less than a millimeter. To determine delay times of multiple objects requires estimating the sequence of delay separations by extracting temporal changes in the interference pattern of the echoes. The models have been previously proposed to determine delay times of multiple objects by using the frequency modulation sound. In order to extract the temporal changes, Gaussian chirplets with a carrier frequency compatible with emission sweep rates were used and each delay time was estimated by using the temporal changes. It was examined that this model could estimate each delay time within the accuracy less than 1 mm in the case of measuring echoes from fixed objects. In this paper, the echoes were measured from moving objects, and it was examined that this model could estimate each delay times of these objects.

3pAB3. Auditory brainstem response in the lesser scaup (*Aythya affinis*), a species of diving duck. Sara C. Therrien, Catherine E. Carr (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, therrien@umd.edu), Elizabeth F. Brittan-Powell (Univ. of Maryland, College Park, MD 20742), and Alicia M. Wells-Berlin (Patuxent Wildlife Ctr., Laurel, MD 20708)

The auditory brainstem response (ABR) is an effective approach to describing auditory sensitivity in a variety of animals. In this investigation, we used the ABR to estimate the auditory sensitivity of a species of diving duck, the lesser scaup (*Aythya affinis*). The typical lesser scaup ABR waveform showed two to three prominent peaks that occurred within the first 5 ms after onset of the stimulus. Peak amplitude increased and peak latency decreased with increasing stimulus sound pressure level (SPL). Threshold was defined as 2.5 dB below the lowest SPL that evoked a visual response (visual detection method). The best range of hearing for this species was from 500 to 5000 Hz, with sensitivity peaking between 1500 and 3000 Hz. Both the waveform morphology and response characteristics of the peaks to changing stimulus intensity are similar to those found in other avian species, such as screech owls (*Megascops asio*) and budgerigars (*Melopsittacus undulatus*).

3pAB4. Texture discrimination by echolocation in the big brown bat. Ben Falk (Neurosci. and Cognit. Sci. 1147 Bio/Psych Bldg., Univ. of MD, College Park, MD 20742, bfalk@umd.edu), Murat Aytekin, and Cynthia, F. Moss (Univ. of Maryland, College Park, MD 20742)

Echolocating bats emit high-frequency vocalizations and listen to the returning echoes to sense their environment. Echolocation transforms the physical environment into acoustic energy which bats use for navigation and prey capture. This study examined the capabilities of the big brown bat, *Eptesicus fuscus*, to identify and discriminate textured objects. The bat was trained to discriminate a smooth object from a textured object, both spherical, ~ 15 mm diameter, and tethered to the ceiling in a large flight room. The bat was free to fly and investigate the objects before making its decision (signaled by tapping one of the objects in flight − correct=smooth). The bat's discrimination performance was measured with a series of textured objects. Quantification of texture was made through acoustic recordings of the objects. High-speed video recordings allowed a reconstruction of the bat's three-dimensional flight path, full bandwidth ultrasound-sensitive microphones allowed recordings of the bat's sonar emissions, and an array of microphones positioned around the flight room allowed a reconstruction of the bat's sonar beam pattern. The relationship between the bat's performance and the acoustic recordings of the stimuli was examined. How the bat's motor behaviors contribute to its discrimination performance was investigated. [Work supported by NSF(REU), NIMH(R01MH056366.]

3pAB5. Repertoire and possible functions of social calls in foraging big brown bats (*Eptesicus fuscus*). Genevieve S. Wright (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, gspanjer@umd.edu), Chen Chiu, and Cynthia F. Moss (Univ. of Maryland, College Park, MD 20742)

While the echolocation calls of big brown bats (*Eptesicus fuscus*) have been studied extensively, this species' communicative vocalizations have received little attention. Calls were recorded from pairs of bats flying in a laboratory flight room in a foraging context. Pairs consisted of naive bats flown with individuals that were either experienced about how to take a tethered insect or naive or of two experienced bats flying together. This situation elicited a variety of vocalizations with properties different from typical *E. fuscus* echolocation calls. Seven different call shapes that occurred only

when two bats were present were identified and categorized. These included chirp-like calls, rising frequency calls, calls that were longer than 20 ms in duration, and series of ultrasonic, chirp-like calls (or "songs") produced only by male bats and displaying individual variation. Some call types were only recorded when a juvenile was present or very rarely in adult-only pairs. Using high-speed video data collected in synchrony with the audio recordings, the contexts and possible functions of various calls are evaluated. The rate at which different call types were recorded with regard to bats' knowledge of the foraging task and the age and sex of bats flown together is also considered.

3pAB6. Behavioral sensitivity to interaural difference cues in the budgerigar (*Melopsittacus undulatus*). Thomas E. Welch, Kristen Garcia, and Micheal L. Dent (Dept. of Psych., Univeristy at Buffalo, SUNY, Buffalo, NY 14260)

Interaural time and level differences cues (ITDs and ILDs) allow organisms to localize sounds and detect and discriminate signals in noise. Behavioral sensitivity to these important cues has only been studied in humans and a few species of animals. Although the abilities of budgerigars to localize, detect, and discriminate a variety of sounds in quiet or noisy conditions are well known, this is the first study to equip birds with headphones so that their sensitivity to ITDs and ILDs can be measured using lateralization procedures. In one experiment, thresholds are measured behaviorally in budgerigars using pure tones (0.5–8 kHz) and broad-band noise as stimuli in a two-alternative (left-right) identification task. In another experiment, sensitivity to changes in ITDs and ILDs of binaural sounds within a repeating background (discrimination task) is measured using the same stimuli as the first experiment. Budgerigars appear to be less sensitive than humans to both ITDs and ILDs, but similar to cats and monkeys. Overall, these lateralization results are generally consistent with the free-field localization abilities of these birds, but their relatively acute sensitivity to interaural difference cues adds support to the idea that budgerigars may be using their connected interaural pathways to enhance directional hearing.

3pAB7. The effects of frequency content on auditory streaming in budgerigars and zebra finches. Kristen A. Garcia (Dept. of Psych., Univ. at Buffalo-SUNY, Buffalo, NY 14260, kgarcia@buffalo.edu), Siddharth Rajaram, Kamal Sen, Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215), and Micheal L. Dent (Univ. at Buffalo-SUNY, Buffalo, NY 14260)

Auditory streaming is a phenomenon that has been documented in a wide variety of animal species. Recently, space, intensity, time, and spectral composition were all found to be important factors for auditory streaming of birdsong by budgerigars and zebra finches, although the cues varied in importance. Those experiments were extended here to further examine the role of frequency characteristics on the auditory streaming of birdsong. The birds were initially trained using operant conditioning procedures to differentially peck keys in response to either a synthetic zebra finch song consisting of five syllables (whole song) or to the same song with the fourth syllable omitted (broken song). Correct responses were reinforced with millet, and incorrect responses were punished with a lights-off timeout period. Once the birds reached high-performance levels with the training stimuli, probe trials were inserted on a small portion of all trials. The probe stimuli contained either a white noise burst in the missing syllable's location or narrowband pieces of the original fourth syllable. Results show that the different cues are differentially effective in eliciting streaming of birdsong by zebra finches and budgerigars, similar to the previous experiments and results from human speech experiments.

3pAB8. Vocal behavior of North Atlantic humpback whales during reunion events on Stellwagen Bank. Kaitlin Palmer (Graduate Program in Acoust., 201 Appl. Sci. Bldg., University Park, PA 16802, kjp205@psu.edu), Susan E. Parks (Penn State Univ., State College, PA 16804-0030), David Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066), Mason Weinrich , Jennifer Tackaberry (Whale Ctr. of New England, Gloucester MA 01931-0159), Alessandro Bocconcelli (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543-1050), and Ari Friedlaender (Duke Univ., Beaufort, NC 28516)

Humpback whales on their feeding grounds are typically found in small, ephemeral groups which have been hypothesized to form and separate to maximize foraging efficiency, but little is known about the functional mechanisms by which these groups change their composition. In this study, we test the hypothesis that specific vocalization types are associated with events where two or more individual humpback whales join to form a cohesive group on their summer feeding grounds. In the summer of 2008, 11 Dtags (a combined acoustic and movement recording tag) were attached to individual humpback whales on their feeding grounds in the Stellwagen Bank National Marine Sanctuary. Behavioral observations were collected from tagged animals at the surface and the timing of join events were recorded. Calls from the tagged whale and nearby associates were identified and analyzed to test for call rate changes between 10 min periods around joining events and 10 min control periods. Additional analyses included call type usage during joining events. Our results indicate a trend toward increased call rates during join events when compared to randomly selected control periods, with evidence of differential usage of specific call types during join events.

3pAB9. Measurement of Copepodid and Nauplii populations in the Golfo della Biodola using inverse acoustic methods. John Fornshell (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, La Spezia 19126, Italy)

Acoustic observations and simultaneous collection of plankton were made in the Gulf of Biodola on the northern coast of the Island of Elba, Italy. The fieldwork was conducted on the 10th and 11th of October 2009 on the R/V LEONARDO. The acoustic measurements were made using the autonomous underwater vehicle, MUSCLE, at a working frequency of 300 kHz. Plankton samples were also collected simultaneously with a 1/4 m Nansen net with a 20 mm mesh size. The acoustic data were processed by scientists at the NATO Undersea Research Center and provided to the author. The waters of the southern Ligurian sea on this coast are exceptionally clear with very low plankton population densities. The plankton were sorted, identified, and enumerated at the National Museum of Natural History, Smithsonian Institution. It is assumed in this study that the volume back-scatter can be attributed to the plankton in the water column. The numbers in the net samples are compared with the values from the inverse acoustic methods.

3pAB10. Discovery of sound in the sea: An on-line resource. Kathleen J. Vigness-Raposa (Marine Acoust., Inc., 809 Aquidneck Ave., Middletown, RI 02842, kathleen.vigness@marineacoustics.com), Gail Scowcroft, Christopher Knowlton, and Holly Morin (Univ. of Rhode Island, Narragansett, RI 02882)

The scientific community and the public have become increasingly aware of, and concerned about, underwater sound. There is increasing interest in learning about sources and uses of sound, and potential effects of sound on the environment. Underlying this interest, however, is a need to provide scientific information at a level appropriate for the general public and for educational and media professionals. The discovery of sound in the sea website (http://www.dosits.org) provides scientific content introducing the physical science of underwater sound and how people and animals use sound to accomplish various tasks. It also includes three major resource sections. The Media Resources include a Facts & Myths quiz, Frequently Asked Questions, and PDF reprints of a tri-fold pamphlet and a 12-page educational brochure. The Teacher and Student Resources include structured tutorials and educational games. DOSITS recently launched a significantly redesigned site. The look and feel have been refreshed without losing functionality and content. The redesign includes an interactive front page, an interactive Audio Gallery, and a redesigned Scientist Gallery. DOSITS provides easy, efficient access to timely information on the science of underwater sound and the current state of knowledge of the effects of underwater sound on marine mammals and fishes.

3pAB11. Lake bottom forward scattering measurement and model/data comparison. Jinrong Wu and Li Ma (Inst. of Acoust., Chinese Acad. of Sci., No. 21, BeiSiHuan XiLu, Beijing, 100190, China, ymwjr@yahoo.com.cn)

The lake bottom scattering function depends, in general, on the grazing angles and the azimuthal angles of the incident and scattered energy. However, most measurements were for backscatter only. Little was understood about forward bottom scattering in shallow water. In order to address this issue, an experiment was conducted in Qiandao Lake using 16 element ver-

tical array and 8 element horizontal array. The measurement result indicated that there was relatively strong scattering in the forward direction near the specularly reflected ray, while the scattering was much weaker and relatively isotropic away from the specular ray. The Ellis and Crowe's three-dimensional scattering function was used to predict the forward scattered energy. Comparsion between model and measured data has shown that the model can explain the lake bottom scattering well, and the frequency dependence of bottom scattering function from 5 to 9 kHz can be neglected. [Work supported by the NSFC.]

3pAB12. Analyzing acoustic scenes in reverberant environments: A bird study. Sandra H. Blumenrath and Robert J. Dooling (Dept. of Biology & Psych., Univ. of Maryland, College Park, MD, shb@umd.edu)

In reverberant environments, the auditory system processes complex signals that are distorted and overlapped by tails of echoes of preceding sounds. These reverberation effects can be used to discriminate between different types of acoustic environments and to locate a sound source, but it also limits the transfer of biologically important information. To examine the positive impact of reverberation on acoustic scene analysis, we used psychophysical methods to test the ability of canaries, zebra finches, and budgerigars to discriminate between reverberation differences. Negative effects were investigated by asking whether birds can segregate reverberated songs from multiple individuals despite degraded auditory grouping cues. Birds were able to hear small changes in reverberation, although this ability was somewhat dependent on both the species and, in particular, the type of signal. Species-specific vocalizations as well as broadband signals posed the smallest difficulty. Interestingly, the non-oscine, flock-living budgerigar outperformed the other two oscine species. Moreover, birds were less able to "hear out" a target song among a cacophony of other songs when all vocalizations were reverberated. The results suggest both positive and negative impact of reverberant environments on the perception of vocal signals and the analysis of complex acoustic scenes in the tested species.

WEDNESDAY AFTERNOON, 21 APRIL 2010 GRAND BALLROOM V, 1:00 TO 2:00 P.M.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Michelle E. Swearingen, Chair
Construction Engineering Research Lab., P.O. Box 9005, Champaign, IL 61826-9005

Invited Papers

1:00

3pID1. Data-based speech synthesis: Progress and challenges. H. Timothy Bunnell (Ctr. for Pediatric Auditory and Speech Sci., Nemours Biomedical Res., 1600 Rockland Rd., Wilmington, DE 19803, bunnell@asel.udel.edu) and Ann K. Syrdal (AT&T Labs.-Res., Florham Park, NJ 07932)

In contrast to knowledge-based techniques such as articulatory or rule-based formant synthesis, data-based approaches to speech synthesis derive synthetic voices from a corpus of recorded human speech. Current "unit selection" data-based systems can produce synthetic utterances that rival human speech in intelligibility and naturalness by concatenating units of various sizes derived from large searchable corpora of speech aligned with tags that capture a rich set of segmental and prosodic features assigned to phonemic or subphonemic units. Emerging data-based techniques use machine learning algorithms to derive parameters automatically for parametric speech synthesis. Whatever the approach, because the extensive corpus of speech data used to construct most modern data-based systems is obtained from an individual talker, the synthetic output of the system captures much of the vocal identity of that individual. This latter feature has opened up new areas of application, such as personalized synthetic voices and "voice banking" for patients who might lose the ability to speak. This talk will trace the development of current data-based approaches, describe applications of this synthesis technology, and discuss some of the major challenges driving current research efforts. [Some work supported by NIDCD Grant R42-DC006193.]

1:20

3pID2. Hot topics in structural acoustics. Dean E. Capone (Appl. Res. Lab., Penn State Univ., State College PA 16804) and Matthew S. Allen (Univ. of Wisconsin-Madison)

In this paper, an overview of the many opportunities and new research focus areas that are hot topics within the structural acoustics community in the Society will be presented. The talk will discuss the topics within the general categories of computational, analytical, and experimental structural acoustics. In particular, numerical methods in structural acoustics are receiving considerable attention, and a variety of proposed new approaches to structural acoustic computational modeling will be reviewed.

1:40

3pID3. Hot topics in musical acoustics. Thomas D. Rossing (CCRMA, Stanford Univ., Stanford, CA 94305, rossing@ccrma.stanford.edu) and James P. Cottingham (Coe College, Cedar Rapids, IA 52404)

Musical acoustics deals with the production, transmission, and perception of musical sound, as well as the scientific study of musical performance. It interfaces with physics, music, psychology, engineering, architecture, audio engineering, and neurology, and it is often at these interfaces that some of the hottest topics occur. For example, functional magnetic resonance imaging (fMRI) and electro- and magneto-encephalography (MEG and EEG) have made it possible to study music and the brain and to localize neural processing of musical activity. Digital computers have made it possible to synthesize musical sound with every increasing refinement using such

methods as physical modeling, waveguide models, and commuted synthesis. Systems for real-time, high-quality, and low-latency audio over the internet have made it possible for performers at distant locations to play in concert. New acoustical musical instruments, as well as modifications of existing instruments, enrich musical performance, while new performing groups such as laptop and cellphone orchestras appear at many universities.

WEDNESDAY AFTERNOON, 21 APRIL 2010 LAUREL A/B, 1:00 TO 2:00 P.M.

Session 3pNS

Noise, Physical Acoustics, and INCE: Military Noise Environments: Continuous and Impulsive II

Richard L. McKinley, Chair
Air Force Research Lab., 2610 Seventh St., Wright Patterson Air Force Base, OH 45433-7901

Contributed Papers

1:00

3pNS1. Speech intelligibility of air conducted and bone conducted speech over radio transmission. Phuong Tran and Thomas Letowski (U.S. Army Res. Lab., 520 Mulberry Point Rd., Aberdeen Proving Ground, MD 21005, ptran@arl.army.mil)

Personal radio-communication devices are traditionally used with built-in or external air conduction microphones. Recently, there have been some efforts to use bone conduction microphones mounted on the operator's head as a replacement for external air microphones. The goal of this study was to compare the intelligibility of speech transmitted through a boom-type noise-cancelling air microphone and a bone microphone using Thales multiband inter-intra-team radios. Eight participants were divided into four pairs with each participant serving as both a talker and a listener. The Callsign Acquisition Test was used to collect speech intelligibility data with the talker in quiet and in 100 dB(A) noise. The listener was always located in a quiet environment. Both types of microphones resulted in speech intelligibility scores ranging from 100% to 94% in quiet and from 100% to 90% in noise condition. There was no statistically significant difference in speech intelligibility between the two types of microphones. The advantages and disadvantages of using bone microphones under various civilian and military conditions are discussed.

1:15

3pNS2. Glass breakage by blasts: A prediction model and real-world validation. Louis C. Sutherland (LCS-Acoust., 27803 Longhill Dr., Rancho Palos Verdes, CA 90275-3908)

The structural elements most sensitive to blast damage are usually windows. A classic paper by Hershey *et al.* [J. Acoust. Soc. Am. **55**, 1009–1017 (1974)] developed a powerful statistical model for damage to windows from impulsive (sonic boom) pressure loads. The key part of the model depended on both the pressure load and failure stress having a log normal distribution; that is, the log of these variables was normally distributed. This basic concept, applied by the author in several studies on blast and sonic boom damage, is briefly reviewed here for application to blast damage to windows. Validation of the model is provided by data from an accidental explosion of an equivalent 20 lbs of TNT near a nearby nine-building apartment complex. 58 windows were broken, fortunately without any injuries. The predicted damage was augmented by results of a model study of diffraction around a cubical obstacle to approximate the changes in the net pressure loading on the apartment building windows located at various orientations to the blast wave front. The predicted window breakage rate, varying from 0.5% to 61% compared well with the observed values in the range of 0–64%. [Work supported by Wyle Labs.]

1:30

3pNS3. Assessing attitudes toward military blast noise. Edward T. Nykaza (Construction Eng. Res. Lab., U.S. Army Corps of Engineers, 2902 Newmark Dr., Champaign, IL 61822, edward.t.nykaza@usace.army.mil), Kathleen Hodgdon (Penn State Univ., State College, PA 16804-0030), George Luz (Luz and Assoc., Baltimore, MD 21212), and Peg Krecker (PA Consulting Group, Madison, WI 53719)

Existing impact assessment procedures for large weapon noise or blast noise do not fully meet the military's noise management needs. Military noise impacts currently are assessed in terms of the response metric annoyance as predicted by a long-term average noise level metric. This method has proven to be unsatisfactory for impulsive military noise. As a result, ERDC-CERL and SERDP have launched a series of field studies to enhance the understanding of human response to military blast noise. An overview of the four research protocols employed in this project will be presented, and the most recent findings from a qualitative personal interview study and a complaint and annoyance comparison study will be given. The preliminary results indicate that residents living near military installations adapt to the basic noise environment over time, notice unusually large noise events or events that produce house vibrations, and indicate that those who file complaints are significantly more annoyed on average than their neighbors who were exposed to a similar noise environment. All findings in this project thus far indicate that the number, timing, and level of discrete blast events are important for predicting the human response to blast noise.

1:45

3pNS4. Analysis of animal blast overpressure study data to compare current impulse noise criteria and propose some new criteria. Steve Goley, Wonjoon Song, Jay Kim (Dept. of Mech. Eng., Univ. of Cincinnati, Cincinnati, OH 45221-0072, steve.goley@gmail.com), and William J. Murphy (CDC/NIOSH Hearing Loss Prevention Team, Cincinnati, OH 45226, wjm4@cdc.gov)

The effects of high-intensity impulsive noise on the auditory system are complex and difficult to quantify in a simple criterion. Several impulse noise criteria (e.g., MIL-STD 1474D, Pfander, Smoorenburg, LAeq8hr, and a model-based method called AHAAH) are studied for their analytic structure to understand their underlying basic assumptions. Then, these criteria are compared for their performances based on animal blast overpressure data obtained from 905 chinchillas exposed to 137 exposure conditions. The noise exposures consisted of different impulse sources, peak levels, number of impulses with various temporal spacing, and presentations in reverberant and non-reverberant conditions. The noise data are analyzed using the analytic wavelet transform to study their time-frequency structures in relation to patterns of measured auditory damage of chinchillas. Based on these studies, a few new metrics are proposed by taking advantage of capabilities of the advanced signal analysis tool. The statistical correlations of the new metrics with the chinchillas' observed auditory damage are compared with the MIL-STD 1474D, Pfander, Smoorenburg, LAeq8hr, and the AHAAH model damage risk criteria. Possibility of developing a new impulse noise standard according to the approach adopted in this study as well as remaining challenges are discussed.

Session 3pPA

Physical Acoustics, Structural Acoustics and Vibration, and Noise: Sonic Boom II

Natalia V. Sizov, Cochair
Wyle Laboratories, 241 18th St., Ste. 701, Arlington, VA 22202

Kimberly Lefkowitz, Cochair
CSTI Acoustics, 16155 Park Rd., #150, Houston, TX 77084-5100

Victor W. Sparrow, Cochair
Pennsylvania State Univ., Graduate Program in Acoustics, 201 Applied Science Bldg., University Park, PA 16802

Invited Papers

12:45

3pPA1. Field experiment on vibration and acoustic responses of building to sonic booms using impulsive sound source. Yusuke Naka (Supersonic Transport Team, Japan Aerosp. Exploration Agency, 6-13-1 Osawa, Mitaka, Tokyo 181-0015, Japan, naka.yusuke@jaxa.jp), Tetsuya Doi, and Jiro Kaku (Kobayashi Inst. of Physical Res., Kokubunji, Tokyo 185-0022, Japan)

Annoyance related to sonic boom is believed to be enhanced in indoor spaces by vibration of building walls, window rattles, room acoustic effects, and so forth. For investigating such vibration and acoustic effects of sonic booms in buildings, both laboratory and field experiments have been conducted. Although propagation of sonic booms in space and relative phenomena such as diffraction around a building can be reproduced in field experiments, it is sometimes not straightforward to generate sounds similar to sonic booms in open space. In this study, an easy-to-handle, portable device suitable for simulating sonic booms in field experiments is developed. A device developed by some of the authors to generate impulsive sound with fair level of infrasonic components by using compressed air is modified to simulate the negative overpressure of sonic booms. The improved device is used in a field experiment to investigate vibration and acoustic responses of a small house. Results of this preliminary experiment will be reported.

1:05

3pPA2. Rattle laboratory testing of windows exposed to sonic boom excitation. Natalia V. Sizov and Christopher M. Hobbs (241 18th St., Ste. 701, Arlington, VA)

One of the major factors contributing to the annoyance indoor listeners experiencing sonic booms is window rattle. In earlier work the authors experimentally obtained a data set of responses of a large number of different window types installed in residential houses. Several of the windows studied were removed intact from tested houses and brought to a laboratory facility for further testing. They were affixed to a massive upright stand that simulated the house wall. These windows were acoustically excited using the same portable sonic boom simulator as in the *in-situ* tests. The results collected during the laboratory testing showed that rattle sources depend on window construction. Main rattle mechanisms were identified. Based on the knowledge gathered, a semi-empirical rattle model was developed defining the key parameters. Vibrational response of the window subjected to sonic boom excitation was modeled as a single degree of freedom system. The ratio between window element velocity and rattle-radiated sound pressure was obtained experimentally for the laboratory tested windows. The rattle-radiated sound level in the laboratory was compared with sound pressure levels received in front of house window exposed to typical and to low booms during NASA supersonic flight testing in 2007. [Work supported by NASA.]

1:25

3pPA3. Human response to sonic booms and rattle heard indoors. Alexandra Loubeau, Jacob Klos, Jonathan Rathsam (NASA Langley Res. Ctr., MS 463, Hampton, VA 23681, a.loubeau@nasa.gov), Brenda M. Sullivan (Lockheed Martin, NASA Langley Res. Ctr., Hampton, VA 23681), and Joseph R. Gavin (Gulfstream Aerosp. Corp., Savannah, GA 31408)

Human response to sonic booms heard indoors is affected by the generation of contact-induced rattle noise. The effect of sonic boom-induced rattle noise was studied in a series of psychoacoustics tests. In order to remove the effects of loudness, sounds were equalized to the same perceived level (PL) in each test. Stimuli were divided into three categories and presented in three different studies: isolated rattles at the same calculated PL, sonic booms combined with rattles at the same calculated PL, and sonic booms combined with rattles at three different PL. The sonic booms, both measured and synthesized, were filtered to simulate presentation inside structures with different transmission and reverberation properties. The rattle sounds due to sonic booms or mechanical loading were recorded in a residential home. Subjects listened to sounds over headphones and were asked to judge the level of annoyance of other factors, depending on the test. Annoyance to different rattles is shown to vary, and the combination of sonic booms and rattles can be more annoying than the sonic boom alone. These results show that the PL metric does not fully represent human annoyance to the selected indoor sonic boom and rattle sounds.

1:45

3pPA4. An examination of repeatability in physiological responses to simulated low amplitude sonic booms of various peak levels and rise times. Andrew Marshall and Patricia Davies (Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031, aimarsha@purdue.edu)

Findings from recent low-level sonic boom studies have shown that people's evaluations of startle could not be explained by loudness only and the startle evaluations were correlated strongly with evaluations of annoyance. A pilot study was conducted to examine how changes in physiological responses (skin conductance, pulse rate, and EMG) compared with subjective ratings startle. Analysis of this data was difficult because the intervals between stimuli were not long enough for some of the physiological measures to return to baseline before the next stimulus exposure. In addition, repeatability of responses was difficult to measure because of the complexity of the test which involved the subject doing several tasks while being exposed to the sounds. To examine what level of physiological response repeatability can be expected in these sonic boom experiments, another test was conducted in which subjects were exposed to simulated sonic booms. The interval required between boom exposures necessary to return to baseline was also examined. To examine the day-to-day repeatability of these measures, each subject completed several sessions on separate days. The variations in physiological responses within and between subjects will be described along with the trends observed related to signal characteristics.

2:05

3pPA5. Assessing sonic boom responses to changes in listening environment, signature type, and testing methodology. Denise M. Miller and Victor W. Sparrow (Graduate Program in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, dmm503@psu.edu)

The aim of this research was to investigate human response to low-amplitude sonic booms, specifically focusing on listening environment, character variation in sonic boom signatures, and subjective test methods. Two tests were conducted in which subjects rated annoyance of numerous low-amplitude sonic booms using magnitude estimation and categorical line scaling methods. One test utilized Gulfstream Aerospace Corporation's Supersonic Acoustic Signature Simulator (SASSII), while the other was conducted in typical office and outdoor patio settings using a headphone playback system. The test stimuli consisted of indoor and outdoor recordings of low-amplitude sonic booms recorded by NASA researchers during flight tests at Edwards Air Force Base. The results include a comparison of subjective methods as well as the effect of listening environment on human annoyance of low-amplitude sonic booms. [Work sponsored by the NSF and the PARTNER Center of Excellence.]

Contributed Papers

2:25

3pPA6. Output-only modal testing of simple residential structures and acoustic cavities using the response to simulated sonic booms and ambient excitation. Joseph Corcoran, Ryan Haac, Marcel Remillieux, Ricardo Burdisso (Dept. of Mech. Eng., Vibs and Acoust. Labs., Virginia Tech, Blacksburg, VA 24061, jmc82@vt.edu), and Georg Reichard (Virginia Tech, Blacksburg, VA 24061)

Structural modal properties of single-room and two-room rectangular structures built with typical residential construction are extracted using only the vibration responses to (1) a sonic boom simulated with a linear distribution of detonating cord and (2) ambient excitation. Then, the acoustic modal properties of the cavities enclosed by the residential structures are extracted using the pressure responses to the same two types of excitation. The technique involves cross-correlating the responses measured at many points to the responses at a few different reference points on the structure or in the cavity. This allows for the calculation of impulse responses at many points around the structure or cavity due to excitation at the reference points. Structural natural frequencies, modal damping ratios, and mode shapes obtained from the output-only analyses are presented and compared to the results from conventional impact hammer modal testing and from a finite element model, demonstrating good agreement. Acoustic natural frequencies and modal damping ratios extracted in the output-only analyses are presented and compared with room pressure spectra measurements made with speaker noise excitation and analytically calculated results for the single room and the two room configurations, also showing good agreement between analyses. [Work supported by NASA Langley.]

2:40

3pPA7. Structural response to low-amplitude sonic booms. Joseph Gavin (Gulfstream Aerosp., 32 Lummus Dr., Savannah, GA 31408, joseph.gavin@gulfstream.com), Jacob Klos, Alexandra Loubeau (NASA Langley Res. Ctr., Hampton, VA., 23681), and Brenda Sullivan (Lockheed Martin, NASA Langley Res. Ctr., Hampton, VA., 23681)

Two highly instrumented homes have been ensonified by low-amplitude sonic booms. The resulting measurements have been analyzed (1) to gain insight to the dominant physical acoustics associated with sonic booms heard indoors and (2) to rank order the mechanisms needed for high-quality auralizations. The mechanisms turn out to be both clear and elegant. Diffraction loading on the exterior surface is as expected with an enhancement around on the incident side of the home and a strong shadow zone at higher non-dimensional frequencies. The resulting pressure loading induces modal vibration of the walls, ceiling, and windows; each of which show their first flexural modes at frequencies below 20 Hz. The walls themselves appear to break up into multi-layered wave bearing structures around 100 Hz, above which the nonlinear contact and sliding mechanisms take on a dominant role in the signature. These insights have been combined to motivate a low-dimensional tool for auralizing the complete audio bandwidth of booms heard indoors.

2:55

3pPA8. Synthesis of sonic boom noise inside structures: An innovative experimental approach. Patricio Ravetta, Ricardo Burdisso, and Nikolas Macko (AVEC, Inc., 3154 State St., Ste. 2230, Blacksburg, VA 24060, pravetta@avec-engineering.com)

Current approaches to evaluate the impact of sonic booms on structures are rather few and limited. In this work, an innovative approach was developed and experimentally validated. It consists in simulating the distributed boom load on the structure using an array of discrete forces. Each discrete point force is the integrated sonic boom pressure loading (from experimental data or a diffraction code) over an area associated with that force. Under the assumption that the structure behaves linearly, each discrete force is applied sequentially and the desired vibro-acoustic responses recorded simultaneously with the applied force. The vibro-acoustic responses due to all forces applied simultaneously (and simulating the sonic boom load) can then be obtained simply as a linear superposition of the individual responses. Since the actual force time history due to the sonic boom does not need to be generated, the approach can be implemented using a shaker or modal hammer. Thus, it is sufficient to measure the transfer function between the input force and the vibro-acoustic response of interest. The vibro-acoustic data are then synthesized using the measured transfer functions and the loading on the structure to estimate the time histories of the forces due to the booms.

3:10

3pPA9. Modification of the nonlinear wave progressive equation model for sonic booms in a stratified atmosphere. B. Edward McDonald (Naval Res. Lab., Washington, DC 20375, mcdonald@sonar.nrl.navy.mil) and Andrew A. Piacsek (Central Washington Univ., Ellensburg, WA 98926)

The nonlinear progressive wave equation (NPE) model [McDonald and Kuperman, J. Acoust. Soc. Am. **81**, 1406 (1987)] is modified to simulate sonic boom physics in a stratified atmosphere using a wave—following window whose dimensions are much less than atmospheric scale heights. (The NPE was originally formulated for time domain weak shock propagation in the ocean, where the ambient density is nearly constant.) This allows simulation of sonic boom propagation in a refracted ray tube in which the ambient density may vary considerably as the simulation window moves along. Energy conservation places a strong constraint on the form of the resulting nonlinear wave equation. Properties of the modified NPE and some illustrative solutions are presented. [Work supported by the Office of Naval Research.]

WEDNESDAY AFTERNOON, 21 APRIL 2010 GRAND BALLROOM III/IV, 1:00 TO 2:45 P.M.

Session 3pPP

Psychological and Physiological Acoustics: Perception of Sounds in the Natural Environment

Robert A. Lutfi, Chair
Univ. of Wisconsin, Dept. of Communicative Disorders, Madison, WI 53706

Contributed Papers

1:00

3pPP1. A method for including manner of contact in the synthesis of impact sounds for perceptual research. Christophe N. J. Stoelinga and Robert A. Lutfi (Dept. of Communicative Disord., Auditory Behavioral Lab., Univ of Wisconsin, Madison, WI 53706)

Impact sounds synthesized according to a physical model have increasingly become the stimulus of choice in studies of sound source perception. Few studies, however, have incorporated manner of contact in their models because of the many parameters entailed in its description. Based on the work of Zener [Phys. Rev 669–673 (1941)], we show that the seven Hertzian parameters required to completely describe the contact can be reduced to just three: force of contact, duration of contact (t_c) and a non-dimensional parameter (λ). Using only these parameters for the contact, a simplified method is presented to synthesize the impact sounds of simply supported plates. The results of the method are shown to be in good agreement with those of the more comprehensive method of Chaigne and Lambourg [J. Acoust. Soc. Am. 1422–1432 (2001)], the only such method to receive validation through acoustic measurement of plates. Psychometric functions obtained from listeners for the discrimination of t_c and λ also show these parameters to be clearly discriminable over the range for which they would realistically vary. Potential advantages of using the simplified method over more comprehensive methods of sound synthesis in perceptual studies are discussed. [Research supported by NIDCD Grant No. 5R01DC006875-05.]

1:15

3pPP2. Acoustic analysis of perceptual sound categories. Laurie M. Heller (Dept. of Psych., Baker 254p, Carnegie Mellon Univ., Pittsburgh, PA 15213, laurieheller@cmu.edu) and Benjamin Skerritt (Brown Univ., Providence, RI 02912)

Acoustic analyses were conducted on several groups of sound events. The sound events were derived from a combination of physical analyses and perceptual data. The acoustic analyses utilized both low-level properties, such as spectral centroid, and higher-level properties. The higher-level properties were based on a combination of the physical source characteristics of basic classes of sound-generating events, statistical methods, and perceptual characteristics. A range of metrics are used to evaluate the success and generality of these acoustic properties, including classification performance and comparison to other results in the literature. (Work supported by NSF.)

1:30

3pPP3. A perceptive categorization of sounds produced the interaction of solid objects. Olivier Houix, Guillaume Lemaitre, Nicolas Misdariis, Patrick Susini (IRCAM, 1 place Stravinsky, 75004 Paris, France), and Isabel Urdapilleta (Universit Paris VIII, F-93526 Saint-Denis cedex, France)

The literature on the perception of environmental sounds often postulates that these sounds are perceptually organized according to their cause. In a previous study, we found experimental results consistent with this assumption by showing that, in the case of identifiable sounds and non-expert listeners, main categories of liquids, gases, electrical devices, and solids appear. We report here a study specifically exploring this latter category. Participants were required to categorize a set of sounds according to the actions causing the sound and to describe the resulting categories. We analyze the classifications with a specific hierarchal cluster analysis producing two dendrograms. The second dendrogram is used to highlight a possible secondary categorization that may add to the primary categorization. Therefore, this technique allows taking into account the multiple types of similarity that might occur in a categorization task. This statistical analysis of the classifications was complemented by a linguistic analysis of the descriptions of the categories (ALCESTE and LEXICO). This analysis shows two main categories of descriptions: those describing several objects in interaction and those describing modifications of an object. Eventually, we discuss how hierarchical cluster and linguistic analyses complement to provide a portrait of the perceptive categorization of sounds of solids.

1:45—2:00 Break

2:00

3pPP4. Signal detection as a function of relative acoustic entropy. Christian E. Stilp (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, cestilp@wisc.edu), Michael Kiefte (Dalhousie Univ., Halifax, NS B3H 1R2, Canada), and Keith R. Kluender (Univ. of Wisconsin, Madison, WI 53706)

Natural sounds are far from random. Instead, they possess structure that reflects physical constraints on sound-producing objects and events. Listeners likely exploit such structure (predictability and redundancy) when perceiving complex natural sounds. The present experiments investigate listeners' ability to discover and use structure in novel sounds to identify a repeating target sound against a competing background of sounds. Complementary signal-processing strategies independently varied acoustic entropy (complement to redundancy) in frequency or time. Novel sounds were generated by stretching the entropy in a restricted frequency band (instantaneous frequency dilation) or temporal slice (phase vocoding) of pink (1/f) noise across the bandwidth or duration, respectively, of an unstretched noise sample. Listeners matched a probe sound to a target sound (1%, 3.2%, or 10% entropy of pink noise) that repeats amidst distractor sounds (1%, 10%, or 100% entropy) at 0-dB SNR. In separate experiments varying entropy in frequency or time, identification of target sounds de-

pended critically on the difference in entropy across targets and distractors. Listeners were more proficient identifying lower-entropy targets against higher-entropy distractors but not lower-entropy distractors. Similarly, higher-entropy targets were better identified amidst lower-entropy distractors. Potential implications for cochlear implant processing will be discussed. [Work supported by NIDCD.]

2:15

3pPP5. How do listeners react to different urban soundscapes? A functional magnetic resonance imaging study of perception and emotion. Deborah Hall (School of Psych., Nottingham Trent Univ., Burton St., Nottingham, United Kingdom, deborah.hall@ntu.ac.uk), Amy Irwin (MRC Inst. of Hearing Res., Univ. Park, Nottingham, United Kingdom), and Christopher Plack (Univ. of Manchester, Manchester, United Kingdom)

This research formed one part of a multi-disciplinary positive soundscapes project investigating the perception of urban sound environments. To determine the neural correlates associated with perceptual and affective responses to such sounds, we have measured ongoing brain activity using functional magnetic resonance imaging. Sixteen participants were scanned while passively listening to a set of recordings made in urban spaces. In total, there were 150 recordings, each containing multiple sound sources. Soundscapes were matched in overall sound level [71 dB(A)] but differed in their ratings of pleasantness along a five-point scale. As expected listening to urban soundscapes evoked a significant response in a number of auditory brain regions (inferior colliculus, medial geniculate body, and auditory cortex). Crucially, those soundscapes evoking a strong emotional response (either pleasant or unpleasant) activated one of the main emotional centres in the brain—the amygdala. The perceived pleasantness of the soundscape also modulated the auditory response to the sounds. Our results confirm that that loudness is not the only factor that determines how people react to urban soundscapes.

2:30

3pPP6. Non-intrusive audio notification in emotion classified background music. Ralf Jung (German Res. Ctr. for Artificial Intelligence, Campus D3 2, 66123 Saarbruecken, Germany, ralf.jung@dfki.de)

In human-computer interaction, audio signals are generally used for warnings and acoustic feedback. Especially in environments with many people, such signals often affect intrusively and distractive because of their differentness compared to the natural sound environment. In this paper, a novel audio notification system is presented that provides methods to notify persons individually in multi-user environments. The used audio cues are unnoticeable for others since they can be embedded seamlessly in ambient background music. A novel notification hierarchy with three signal classes is introduced to realize a personalized and a context-aware notification within the background music. For the self-composed background music, perceptual constraints such as auditive Gestalt laws as well as results from musicology were taken into account to influence the listeners perception. A classification model for the emotional expression of non-vocal music is also introduced in this paper. The effectiveness of the novel notification method and the plausibility of the classification model are confirmed by empirical studies. This novel non-intrusive audio notification method let users become aware of individual events while at the same time other present persons will not get distracted by the notification.

WEDNESDAY AFTERNOON, 21 APRIL 2010 DOVER C, 1:25 TO 3:15 P.M.

Session 3pSA

Structural Acoustics and Vibration and Engineering Acoustics: Nondestructive Testing of Materials

Hasson M. Tavossi, Chair

Valdosta State Univ., Dept. of Physics and Astronomy Geosciences, 1500 N. Patterson St., Valdosta, GA 31698

Chair's Introduction—1:25

Invited Papers

1:30

3pSA1. Approaches for ultrasonic damage detection in aircraft components for structural health monitoring. Bernhard R. Tittmann (Dept. of Engr. Sci. & Mech., The Pennsylvania State Univ., 212 Earth-Engr. Sci. Bldg., University Park, PA 16802, brt4@psu.edu)

Characterization of flaws is an important goal of ultrasonic NDE in aircraft components, especially to meet the new demands for structural health monitoring to achieve retirement for cause. This presentation describes approaches and results for two classes of damage: (1) impact damage in fiber-reinforced composites and (2) cracks in jet engine components. In the application of impact damage on composites the use of non-contact laser based ultrasound with a NdYAG laser as transmitter and Mach–Zehnder interferometer as receiver was complemented by acoustic microscopy. Results are described for both hard and soft impacts on a variety of composites. In the application of crack detection of jet engine components the main emphasis was on the development for *in-situ* high-temperature piezoelectric sensors. In particular, high temperature operation in excess of 1000 °C is described with sol-gel spray-on coatings, thin AlN films and single crystals of aluminum nitride. Some of the sensors are geared toward wireless retrieval of crack detection in turbine disks and blades. Whereas these approaches in ultrasonic NDE are not routine they hold high promise for near-future applications in the aerospace.

1:50

3pSA2. Sensor network analysis for fatigue crack characterization in the vicinity of fastener holes. Cliff J. Lissenden, Hwanjeong Cho, and Chung Seok Kim (Eng. Sci. and Mech., Penn State, University Park, PA 16802, lissenden@psu.edu)

A network of surface mounted piezoelectric disk transducers is used to detect, locate, and eventually size a growing fatigue crack in an aluminum plate with multiple fastener holes. The results will be used as input to a probabilistic fatigue crack growth model in order to predict structural reliability and remaining useful life. The target application is built-up aircraft wing and fuselage structures. Since the transducers are bonded to the surface and available upon demand, this is structural health monitoring. A tomography algorithm based on *a priori* knowledge that a fatigue crack will initiate near to the centerline of the fastener holes is used to determine at which hole the crack initiated using Lamb wave modes. In that vicinity, both through-transmission and pulse-echo modes are then used to interrogate many guided wave propagation paths. Use of both modes enables evaluation of transmission and reflection coefficients. Training data are acquired from simpler geometries to facilitate characterization of crack length from received signal features. Characterization of fatigue crack size is in progress. The presentation will conclude with a description of the link to probabilistic modeling. [Support from the Ben Franklin Center of Excellence in Structural Health Monitoring is acknowledged.]

2:10

3pSA3. Ultrasonic nondestructive characterization of inhomogenieties and stress concentrations in the porous materials. Hasson M. Tavossi (Dept. of Phys., Astronomy & Geosciences, Valdosta State Univ., 1500 N Patterson St.,Valdosta, GA 31698.)

Ultrasonic nondestructive bulk testing of porous materials is difficult to achieve due to the high attenuation of ultrasonic waves in these materials. It is shown experimentally that ultrasonic wave attenuation and pass-band frequencies in the porous materials are some functions of average pore size, when other material parameters are kept constant. In most theoretical effective-media models for the porous materials, pore ratio is included but not their average size. The average pore size has a significant effect on both wave attenuation and pass-band frequencies of the porous materials. The goal of this study is to determine optimal criteria for nondestructive flaw detection in the porous materials. Experimental results are obtained on ultrasonic wave attenuation for different samples, as a function of frequency and the average pore size. Flaws in the form of inhomogenieties and localized stress concentrations are then introduced in the samples and the NDT techniques are applied to detect them. Experiments are performed to identify the type of flaw, size, and location in the porous samples. Data on location and sizing of the inhomogenieties and stress concentration will be presented. Results of this study can be used to determine the most effective ultrasonic NDT techniques for flaw detection in the porous materials.

2:30—2:45 Break

Contributed Papers

2:45

3pSA4. Locating a small change in a multiple scattering environment (LOCADIF): Application to monitoring concrete. Thomas Planes, Eric Larose (LGIT-CNRS, BP 53, 38041 GRENOBLE Cedex 9, France, thomas.planes@ujf-grenoble.fr), Vincent Rossetto (LPMMC, CNRS, Grenoble, France), and Ludovic Margerin (CEREGE, CNRS, Aix En Provence, France)

Most conventional imaging techniques are operated in single scattering media where the Born approximation is found relevant. In the multiple scattering regime where this approximation does not hold, the wave forgets its initial direction after a few scattering events and we cannot have access to the precise trajectory of the wave. All waves are mixed together in phase, amplitude, and wavevector, which defines the diffuse regime. This paper presents an imaging technique to locate a weak perturbation in a multiple scattering environment. We derive a formula to predict the spatio-temporal decorrelation of diffuse coda waves induced by an extra-scatterer. Locating this new defect is then formulated as an inverse problem which is solved by a maximum likelihood approach. Using elastic waves in the 100–350-kHz frequency band, we show that the position of a millimetric hole drilled in a heterogeneous concrete sample can be retrieved with a precision of a few cm. We illustrate the robustness of the method and define an optimal time of propagation that depends on the geometry of the acquisition and the diffusion constant of the medium. Applications to monitoring civil engineering structure are discussed.

3:00

3pSA5. Semi-analytical study of wave propagation in wire waveguides. Manton J. Guers and Bernhard R. Tittmann (Dept. of Engr. Sci. & Mech., The Pennsylvania State Univ., 212 Earth-Engr. Sci. Bldg., University Park, PA 16802, mjg244@psu.edu)

Guided ultrasonic waves in wire waveguides are useful for various applications including position monitoring and *in-situ* measurements of specimens in harsh environments. In corrosive environments, it has been observed that the velocity of the guided waves changes due to the growth of an oxide layer on the waveguide. Previous work by Grossman [IEEE Trans. Sonics Ultrason. **31**, 25–31 (1984)] analyzed this phenomenon by calculating the change in rod velocity due to oxide thickness. In this paper, the semi-analytical finite element (SAFE) method was used to determine how the velocity changes with respect to oxide thickness and frequency. In addition, the effects of protective coatings on the waveguide have been considered. Results demonstrate that the SAFE method is effective for determining the dispersion characteristics of a waveguide comprised of two distinct materials.

Session 3pSC

Speech Communication: Speech Intelligibility (Poster Session)

Sarah Hargus Ferguson, Chair
Univ. of Kansas, Dept. of Speech Language Hearing Science Disorders, 1000 Sunnyside Ave., Lawrence, KS 66045

Contributed Papers

To allow for extended viewing all posters will be on display from 9:00 a.m. to 12:00 noon and 1:00 p.m. to 3:00 p.m. All authors will be at their posters from 1:00 p.m. to 3:00 p.m.

3pSC1. Coherent versus incoherent modulation filtering. Marc A. Brennan (Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, mabrenn1@u.washington.edu), Pamela E. Souza (Northwestern Univ., Evanston, IL, 60208), Les E. Atlas, and Adam R. Greenhall (Univ. of Washington, Seattle, WA 98195)

Understanding how amplitude modulations are used for speech recognition is essential for improving speech signal processing techniques. Previous studies using vocoding or incoherent modulation filtering indicated that slow amplitude modulations are essential for speech recognition [e.g., Drullman (1994), Shannon (1995)]. Each method had limitations that constrain our understanding of how speech is decoded. These methods did not account for a product model which has the modulating signal multiply the carrier signal in speech or used modulation filtering which retained inherent carrier components. Coherent modulation filtering [C. & Atlas (2009)] is a modulation filtering method that uses a product model with more effective modulation filtering. The purpose of this study was to establish differences in speech recognition performance with coherent versus incoherent modulation filtering and assess the importance of slow amplitude modulations in speech recognition while preserving the carrier signal. Subjects were adults with normal hearing. Consonant recognition scores were measured for 16 nonsense syllables at different modulation filter cutoff frequencies. Speech was presented at a conversational level in a closed-set task. Recognition improved at high modulation filter cutoff frequencies for both conditions. At low-modulation filter cutoff frequencies, performance was substantially poorer with coherent compared to incoherent modulation filtering.

3pSC2. Age effects in the understanding of noisy speech denoised by estimated ideal binary masks. Pierre Divenyi, Nicholas Livingston (VA Northern California Health Care System and East Bay and Inst. for Res. and Education, Martinez, CA 94553, pdivenyi@ebire.org), Adam Lammert (Univ. of Southern California, Los Angeles, CA 90089), Ke Hu, and DeLiang Wang (The Ohio State Univ., Columbus, OH 43210)

An algorithm was developed to denoise speech presented in matching speech-spectrum noise by estimating the ideal binary mask (IBM) that retains the time-frequency (T-F) regions with favorable local signal-to-noise ratio (SNR) and removes the remaining regions [Hu *et al.*, in Proceedings of the 4th SAPA Workshop, Interspeech 2008]. With this method sentence intelligibility was tested in 20 elderly and 20 normal-hearing young listeners. The elderly group had, at worst, moderate SNHL between 0.5 and 4 kHz. Speech material consisted of IEEE sentences spoken by a single female talker. Intelligibility was tested at two SNRs in the elderly (-2 and -4 dB) and three SNRs in the young group (-4, -6, and -8 dB). Subjects listened to 50 sentences at each SNR in three conditions: denoising by the estimated IBM by the algorithm, denoising by the IBM calculated from premixed signals, and the sentences presented in the original noise without processing. With this last condition taken as the baseline, both keyword- and phonemic intelligibility of the sentences processed with the algorithm showed a slight improvement for the young but no improvement for the elderly listeners, while both groups performed substantially better than the baseline in the IBM condition. [Work supported by the VA Medical Research.]

3pSC3. Automatic scoring of responses to the hearing in noise test using utterance verification based on hidden Markov models. Jason Lilley (Dept. of Linguist., Univ. of Delaware, 46 E. Delaware Ave., Newark, DE 19711, lilley@asel.udel.edu), Justin M. Aronoff, Sigfrid Soli (House Ear Inst., Los Angeles, CA 90057), Tim Bunnell (Nemours A.I. duPont Hospital for Children, Wilmington, DE 19803), and Ivan Pal (Compreval, Inc., Grayslake, IL 60030)

The hearing in noise test (HINT) is widely used in clinical settings to measure patients' speech recognition thresholds (SRTs). The patient must repeat 20 sentences presented at signal-to-noise ratios (SNRs) that change adaptively throughout the test. HINT varies the SNR of the sentences according to the patient's responses, so each response must be scored by a human judge before the next sentence is presented. An utterance verification engine (UVE) would remove both the need for a human judge and the subjectivity of human judgments. 2000 HINT responses were used from 25 listeners to develop a UVE based on hidden Markov models (HMMs), a lexicon of observed and likely words, and a grammar that generates utterances classifiable as correct or incorrect. An evaluation study is in progress with a planned enrollment of 25 normal-hearing listeners. HMMs, lexicon, and grammar are updated after every five listeners. In results with ten listeners tested to date, scoring accuracy averaged 88% per subject (range 81%–95%). The mean SRT determined by the UVE differed from the mean SRT determined by human scorers by an average of 0.56 dB (range 0.04–1.91). Complete results will be presented at the conference. [Work supported by NIDCD Grant R43DC008212.]

3pSC4. Effects of semantic and syntactic cues in (non) native speech perception in noise. Bin Li and Ratree Wayland (Linguist. Program, Univ. of Florida)

Non-native listeners (NNLs) are reported to be less tolerant with the degradation of speech (e.g., speech in noise) than native speakers (NSs). In addition, benefits from clear speech in such conditions have been shown to be greater for NSs than for NNLs in speech perception. This study aimed to examine whether poor perception by NNLs is due to their limited usage of the contextual information or their abilities in processing acoustic-phonetic signals, by investigating Chinese listeners' ability in perceiving English sentences in varying linguistic contexts, such as semantic and syntactic cues, under different listening conditions. Sentences differing in degrees of semantic or syntactic acceptability produced by NSs of English were judged a different group of naive NSs of English on degrees of acceptability. All sentences were digitized and masked with white and babble noises with different signal-to-noise ratios (-4 and -8 dB). Listeners from three language backgrounds were recruited. Results showed that semantic and syntactic cues, as well as listening conditions, exhibited different degrees of facilitation in rewriting of keywords. Chinese speakers benefited more from marked syntactic structures, whereas English speakers found marked semantic contexts more facilitating.

3pSC5. Intelligibility of accented American English isolated vowel: Effects of fundamental frequency (F0) contour and formant structures. Sangeeta Kamdar, Chang Liu, Su-Hyun Jin, and Dasia Gonzales (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, Austin, TX)

Our previous studies have shown that Chinese and Korean accented English vowels deviated in fundamental and formant frequency values and gross spectral shape properties when compared to English-native productions. The present study proposed to examine the intelligibility of accented English vowel by manipulating the fundamental frequency, formant frequencies, and/or entire spectrum shape in an effort to measure the specific contribution of each of these discrete acoustical parameters on accented vowel perception. There were six sets of vowel stimuli: original vowels produced by native and non-native speakers, flat F0 contour with original spectrogram of native and non-native speakers, hybrid vowels mixing non-native F0 contour and native spectrogram, and hybrid vowels mixing native F0 contour and non-native spectrogram. Preliminary results showed that intelligibility was significantly improved for foreign-accented vowels when the non-native spectrogram was replaced with native spectrogram. On the other hand, effects of F0 contour were dependent on talker intelligibility. For non-native speakers with low vowel intelligibility, replacement of F0 contour with native F0 contour slightly increased vowel intelligibility while for non-native speakers with high vowel intelligibility, replacement of F0 contour degraded rather than improved vowel intelligibility.

3pSC6. Speech intelligibility in cross-dialectal multi-talker babble. Robert A. Fox, Ewa Jacewicz, and Chiung-Yun Chang (Dept. of Speech and Hearing Sci., SPA Labs, Ohio State, 1070 Carmack Rd., Columubs OH 43210-1002)

This study examines speech perception in multi-talker babble by varying speaker dialect as a masker. Studies varying the language of the target speech and the babble masker show that the target is most intelligible when the masker is constructed from a non-native language and target represents exemplars from the listener's native language. Does variation in speaker and listener dialect has a similar effect on speech intelligibility? In this study, central Ohio listeners listen to target sentences spoken by either native Ohio or native North Carolina talker. The sentences are masked by multi-talker babble consisting of speech from either Ohio or North Carolina talkers presented at several sound-to-babble (S/B) ratios. The same material is also presented to listeners in North Carolina. The results for Ohio listeners show significant effects of dialect at 0 dB S/B ratio. Intelligibility improved when the background babble featured the dialect that was different from the dialect of the target speaker. The performance of Ohio listeners was more than 20% higher when the target was presented in the North Carolina babble than in the Ohio babble. Results for North Carolina listeners will be discussed. [Work supported by NIH.]

3pSC7. Fricative perception with and without hearing aids, an event-related potential study. Sharon Miller and Yang Zhang (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455)

The present study used auditory event-related potential (ERP) measures to investigate how the adult brain differentially processes fricative speech sounds with and without the use of a hearing aid. Synthetic stimuli for /sa/, /sha/, /as/, and /ash/ were created to control for the spectral cues in the vowel and consonant portions based on naturally recorded speech. ERP responses were recorded for each sound in an unaided and a hearing aid condition using a randomized block design. At least 160 trials per stimulus were averaged for each sound per subject. The results indicated that (1) the ERP responses in the unaided condition significantly differed from the aided condition, (2) N1 peak amplitudes and latencies to /s/ and /sh/ significantly differed in both unaided and aided conditions as well as in syllable-initial and syllable final positions, and (3) phonological context significantly affected N1-P2 responses. Despite some minor differences, these results are consistent with our previous ERP study using natural speech stimuli, which suggests that amplification through hearing aids alters neural coding of acoustic features in speech sounds. [This work was supported by the Graduate Research Partnership Program and Grant-in-Aid, University of Minnesota.]

3pSC8. Suprathreshold auditory processing and speech recognition in noise for hearing-impaired listeners. Van Summers, Joshua G. W. Bernstein, Matthew J. Makashay (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave. NW, Washington, DC), Golbarg Mehraei (Univ. of Maryland, College Park, MD 20742 and Walter Reed Army Medical Ctr., Washington, DC 20307), Sarah Melamed, Marjorie R. Leek, Frederick J. Gallun, and Michelle Molis (VA Medical Ctr., Portland, OR 97201)

Suprathreshold distortions in auditory processing contribute to speech recognition deficits for hearing-impaired (HI) listeners in noise. Outer hair cell damage and attendant reductions in frequency selectivity and peripheral compression may contribute to these deficits. Reduced sensitivity to spectral or temporal modulations or temporal fine structure (TFS) information may also play a role. Eight normal-hearing and 18 HI listeners were tested in psychoacoustic tasks to assess frequency selectivity (notched-noise), peripheral compression (temporal masking curves), TFS sensitivity [frequency modulation (FM) detection in the presence of random amplitude modulation], and spectrotemporal modulation (STM) sensitivity (spectral-temporal "ripple" detection). Performance was examined at 500, 1000, 2000, and 4000 Hz at several presentation levels. Listeners were also tested on sentence recognition in stationary and modulated noise (92-dB sound pressure level signal level; −6-, −3-, 0-, and +3-dB SNRs). HI listeners with similar audiometric thresholds showed a wide range of speech scores. Estimates of FM and STM sensitivity at 500 and 1000 Hz were significantly correlated with speech scores and with each other. Frequency selectivity and compression measures were not clearly associated with speech performance. Implications for models of HI speech intelligibility are discussed. [Work supported by the Oticon Foundation.]

3pSC9. Effect of reduced audibility on masking release for normal- and hard-of-hearing listeners. Peggy B. Nelson, Elizabeth Anderson, Yingjiu Nie, and Bhagyashree Katare (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455, peggynelson@umn.edu)

Listeners with sensorineural hearing loss show reduced benefit from fluctuating compared to stationary maskers. Our hypothesis is that the representation of the signal in the dips is the most important predictor of performance [Jin and Nelson (2006)]. Because the audibility of the signal is the most critical factor, we tested normal-hearing (NH) and hearing-impaired listeners at similar reduced audibility levels. Listeners with normal hearing and hearing loss were presented IEEE sentences at a range of overall levels (from 30- to 80-dB sound pressure level), signal-to-noise ratios, and low-pass filter settings, resulting in a range of signal AIs that varied from 0.1 to 0.95. For the NH listeners, there was a very systematic relationship between AI and performance, but the slope of the function for the higher level stimuli was shallower than that for the low-level stimuli, suggesting that the active mechanism is at work at low levels both in quiet and in noise. The relationship between AI and performance was less systematic for listeners with hearing loss. [Work supported by NIDCD R01-DC008306.]

3pSC10. Relationships between the modified rhyme test and objective metrics of speech intelligibility. Gongqiang Yu, Anthony J. Brammer (Ergonomic Technol. Ctr., Uconn Health Ctr., 263 Farmington Ave., Farmington, CT 06030), Kara Swan, Jennifer B. Tufts (Univ. of Connecticut, Storrs, CT 06269), Martin G. Cherniack, and Donald R. Peterson (Uconn Health Ctr., Farmington, CT 06030)

Relationships between the modified rhyme test (MRT) and the speech transmission index (STI), and MRT and the speech intelligibility index (SII), have been obtained for additive stationary noise and nonlinear distortion of the speech signal. The former was speech-spectrum shaped noise, white noise, or −3-dB/octave noise presented at speech signal-to-noise ratios ranging from −25 to +10 dB. The speech distortions were peak clipping or center clipping, with clipping thresholds from 2% to 98% of the cumulative magnitude histogram. Subjects (4 male, 4 female) with normal hearing were seated in an anechoic chamber. Speech was reproduced by a small, high-fidelity loudspeaker located 2.4 m to the center-of-head and the inter-aural axis. Noise was reproduced by four-loudspeaker systems and processed to simulate a diffuse field at the ear in the horizontal plane. The revised STI

was computed using the standardized test signal [IEC 60268-16 (2003)] and the speech signal. The SII was calculated using coherence to estimate the effects of distortion. Results will be discussed in terms of the ability of each objective metric to predict intelligibility and adjustments to the metrics to improve performance. [Work supported by NIOSH Grant R01OH008669.]

3pSC11. Phoneme perception in temporally and spectrally modulated maskers: Effect of suprathreshold hearing impairment. Sandeep A. Phatak and Ken W. Grant (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC)

Normal-hearing (NH) listeners show better speech recognition in fluctuating maskers than in steady-state (SS) maskers. Hearing-impaired (HI) listeners do not show such benefit due to elevated thresholds. With amplification, HI listeners show some benefit, but significantly less than that for NH listeners. To investigate the reduced benefit in HI listeners at suprathreshold levels, we measured consonant and vowel recognition by NH and aided HI listeners in the presence of SS and modulated maskers. The modulated noises were generated by modulating the SS noise, in either one or six channels, by either one talker speech or four talker babble, resulting in four different modulated masker conditions. HI listeners were provided individualized gain according to the Cambridge formula. Both NH and aided HI listeners showed reduced intelligibility for consonants and vowels due to increase in the number of talkers. Increase in the number of channels provided more benefit to NH listeners than to HI listeners, and this benefit was greater for vowels than for consonants. This benefit due to spectral modulations for aided HI listeners was not correlated with audibility, but seems to be correlated with their ERBs. [Work supported by the Oticon Foundation, Denmark.]

3pSC12. Improvement of consonant identification in noise by spectro-temporal enhancement. Chang Liu (Dept. of Commun. Sci. and Disorder, 1 Univ. Station A1100, The Univ. of Texas at Austin, Austin, TX 78712) and David A. Eddins (Univ. of Rochester, Rochester, NY 14642 and Rochester Inst. of Technol., Rochester, NY 14623)

Our previous studies have shown that perception of vowels and consonants in long-term speech-shaped noise (LTSSN) by listeners with normal and impaired hearing is improved significantly by spectro-temporal enhancement resulting from modifications in the spectral modulation domain. The present study manipulated several parameters of enhancement processing including length of the analysis window and size of fast Fourier transform (FFT) analysis to determine which parameters produce the greatest improvement in stop consonant identification in the presence of LTSSN and cafeteria noise for normal-hearing listeners. The speech signals were presented at 70-dB sound pressure level and the signal-to-noise ratio was varied in fixed steps. Results indicated that identification of stop consonants increased by about 10 percentage points in LTSS noise and by about 25 percentage points in cafeteria noise when the length of analysis window was 200 ms and the FFT size was 512 points. Subsequent acoustic analyses showed that such perceptual improvement may be due to enhancement of formant transitions relative to background competition.

3pSC13. When noise vocoding can improve the intelligibility of sub-critical band speech. James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin-Milwaukee, P.O. Box 413, Garland 224, Milwaukee, WI 53201, bashford@uwm.edu)

This study examined the redundancy of spectral and temporal information in everyday sentences, which were reduced to 16 rectangular spectral bands having center frequencies ranging from 250 to 8000 Hz, spaced at 1/3-octave intervals. High-order filtering eliminated contributions from transition bands, and the widths of the resulting effectively rectangular speech bands were varied from 4% down to 0.5% (Exp. 1) and from 40 Hz down to 5 Hz (Exp. 2). Intelligibility of these sub-critical bandwidth stimuli ranged from nearly perfect in the 4%- and 40-Hz bandwidth conditions, down to nearly zero in the 0.5%- and 5-Hz bandwidth conditions. However, a large recovery of intelligibility was obtained under adverse filtering conditions when the speech bands were used to vocode broader noise bands that approximated critical bandwidths (ERBn) at the 16 center frequencies. For example, the 0.5%- and 1%-bandwidth speech stimuli were only about 1% and 20% intelligible, respectively, whereas scores of about 26% and 60%, respectively, were obtained for the ERBn-wide modulated noise stimuli derived from the speech bands. These large intelligibility increases occurred despite elimination of spectral fine structure and the addition of stochastic fluctuations to the speech-envelope cues. Additional findings and their implications will be discussed. [Work supported by NIH.]

3pSC14. The perceptual organization of noise-vocoded speech under competitive conditions. Brian Roberts, Robert J. Summers (Psych., Sch. of Life and Health Sci., Aston Univ., Birmingham B4 7ET, United Kingdom), and Peter J. Bailey (Univ. of York, Heslington, York YO10 5DD, United Kingdom)

Research using sine-wave speech suggests that across-formant grouping depends critically on modulation of the frequency but not the amplitude contours of the formants. The generality of this finding was explored using noise-vocoded sentences. Sentences were filtered into six frequency bands, and the amplitude envelope of each band was used to modulate a matched noise-band carrier (N). Bands were paired, corresponding to F1 ($=N1+N2$), F2 ($=N3+N4$), and the higher formants ($F3'=N5+N6$), so that the frequency contour of a formant was implied by variation in the relative amplitudes of the corresponding pair of bands. Perceptual organization was probed by presenting stimuli dichotically (F1+F2C; F2+F3′), where F2C is a competitor for F2 that listeners must resist to optimize intelligibility. F2Cs were derived from F2 by time reversing the envelopes of $N3+N4$. Implied frequency variation in F2C was manipulated by varying the AM depth in the constituent bands (100%-0% of original depth) or by comodulating both bands with the same envelope (from $N3$, $N4$, or $N3+N4$). Competitor efficacy declined with AM depth. Comodulated competitors were less efficacious than their 100% modulated counterpart. The findings are consistent with the proposal that it is the time-varying frequency contours of formants that govern their grouping. [Work supported by EPSRC.]

3pSC15. Word perception, intelligibility, and simultaneous task performance. Nirmal Kumar Srinivasan and Thomas D. Carrell (Commun. Disord., Barkley Memorial Ctr., Univ. of Nebraska-Lincoln, Lincoln, NE 68583)

The relationship between speech intelligibility and simultaneous task performance was investigated in the present experiments. Specifically, the following indicators were measured: intelligibility, reaction time for word repetition, visual-motor performance, cognitive load, and visual word recall. These relationships were measured using natural, synthetic, and cell phone speech. It is clear from other reports that different patterns of results may be expected from different types of speech signals [Pisoni *et al.* (1985)]. Intelligibility was based on participant's accuracy in repeating the final word of a sentence. Visual-motor performance was measured as the average speed of rotation in adaptive pursuit rotor task [Srinivasan and Carrell (2007)]. Task load was measured using the NASA-TLX instrument [Hart and Staveland (1988)]. TLX was highly correlated with the simultaneous pursuit rotor task in the cell phone listening condition ($r = 0.41$, $p < 0.0001$). However, the pursuit rotor measure consistently demonstrated much lower variability than the TLX score. Perception was also measured in two attention conditions, consistent-mapping and variable-mapping [Schneider and Schiffrin (1977)]. In the consistent-mapping condition, cell-phone speech led to better performance at the pursuit rotor task than natural speech, whereas in the variable-mapping condition, the natural speech led to better performance.

3pSC16. Linguistic processing of vocoded signals. Susan Nittrouer and Joanna H. Lowenstein (Otolaryngol., Ohio State Univ., 915 Olentangy River Rd., 4th Fl., Columbus, OH 43212, nittrouer.1@osu.edu)

Tests of word discrimination show that listeners can accurately recognize speech with cochlear implants, but there is more to linguistic processing than these measures evaluate. We examined the possibility that other aspects of processing are hindered. Adults and 8-year-olds participated in two experiments. (1) Serial recall for words presented in each of three ways: natural, vocoded, and embedded in noise. Environmental sounds were control stimuli. (2) Judgments of whether words share the same final consonant using natural words, and those same words vocoded or embedded in noise. Serial recall was less accurate and responding was slower for vocoded than for natural speech, but not as poor as for environmental sounds. For final-consonant judgments, accuracy was not as greatly affected by use of vocoded signals, but response times were slowed. That was true even when only correct responses were considered. In both experiments, adults per-

formed better with speech-in-noise than with vocoded speech, but children performed similarly with the two kinds of signals. Conclusions were: as follows (1) linguistic processing of signals heard through implants can be hindered in ways not typically evaluated (2) that hindrance arises from impoverishment of signal properties required for recovering phonetic structure, rather than from general signal degradation. [Work supported by NIDCD Grant No. DC-00633.]

3pSC17. The speech scale. S. Umesh (Indian Inst. of Technol. Madras, Chennai 600 036, India), Leon Cohen (City Univ. of New York, New York, NY 10065), and Douglas Nelson (Dept. of Defence, Fort. Meade, MD, 20755)

Using the Timit database, we describe experimental results that lead to the concept that sounds made by different individuals and perceived to be the same that can be transformed into each other by a speech scale. The speech scale is empirically determined using only speech data. We show the similarity of the speech scale to the MEL scale of Stevens and Volkmann, which was derived only from hearing experiments. We thus argue that we have experimentally linked speech production and hearing.

3pSC18. Native and non-native speech database for children. Tessa Bent (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, tbent@indiana.edu)

Language acquisition, performance in school, and social interactions all require the ability to compensate for the enormous amount of variability in the speech signal. Foreign-accented speech represents one common, real-world source of speech variability, which can be particularly challenging for accurate speech perception. Although maintaining perceptual constancy in the face of variability is necessary for effective communication, little is known about how listeners develop the skills to compensate for speech variability. As a resource for investigating this issue, I am developing a new database of foreign-accented speech with materials appropriate for children. The database includes 28 speakers with 4 from each of the following language backgrounds: American English, Spanish, Mandarin, French, German, Japanese, and Korean. In each language group, half of the talkers are female. Speakers were recorded reading words, sentences, and paragraphs in English, which have been used in previous research and clinical settings with children. The database also includes assessments of objective intelligibility, perceived comprehensibility, and strength of foreign accent by adult native listeners. With this database, I plan to assess the development of the cognitive-linguistic skills that underlie a listener's ability to compensate for variability in the speech signal. [Work supported by NIH-NIDCD R21-DC010027-01.]

3pSC19. A preliminary comparison of sound quality of maternal speech recordings in varied neonatal intensive care unit settings. Andi C. Petito, Jessica Jancek, and Charlene Krueger (Univ. of Florida, 911 NE 6th Ave., Gainesville, FL 32601, adest@ufl.edu)

Since the 1980s, research investigating the beneficial effects of exposure to maternal voice has been steadily increasing. This has been accomplished by playing back recordings of maternal voice to the preterm infants. It is not known, however, how variations in the typical NICU setting alter the sound quality of the original recordings. This is a sub-study to a larger study in which 58 preterm infants will hear a recording of their mother reciting a rhyme twice a day from 28–34 weeks post-menstrual. A pair of microphones will be placed at ear level of the preterm subjects while the original recordings of maternal speech are played. The sound recorded within the incubator will then be run through a sound analyzer (PRAAT) and compared to the same sound qualities of the original recordings. The purpose of this study is to compare the original sound spectrum of a maternal speech recording to spectra obtained while the recording is played within typical NICU settings (open vs closed crib, pillows vs no pillows, etc.) If it is observed that the sound quality is altered due to varied NICU settings, this information will be important in determining NICU sound regulations for future studies employing voice recordings.

3pSC20. Production of sibilant fricatives by children with cochlear implants. Ann E. Todd, Jan R. Edwards, Ruth Y. Litovsky (Univ of WI-Madison, 1500 Highland Ave., Madison, WI 53705, aetodd@wisc.edu), and Fangfang Li (Univ of Lethbridge, Lethbridge AB T1K 3M4, CA)

Children who use cochlear implants (CIs) are likely to show differences in speech production relative to children with normal hearing (NH) due to the suboptimal auditory input provided by CIs. The present study compared word-initial /s/ and /S/ produced by children with bilateral CIs and two comparison groups of children with NH (children matched for hearing age and children matched for chronological age). The first spectral moment (i.e., centroid), a measure which describes the mean frequency differences in the spectra of sibilant fricatives, was calculated at the midpoint of the fricative noise for /s/ and /S/. Results showed that the children with CIs produced less differentiation between /s/ and /S/, had lower centroid frequencies for /s/, and less variability for /s/ than both groups of children with NH. These results may be explained by the perceptual limitations of CIs; /s/ is a sound produced with more energy at higher frequencies, and CIs provide poor frequency resolution for higher frequencies. Currently the amount of coarticulation in productions of /s/ and /S/ with the following vowel is being examined to determine whether the groups show a difference. [Work supported by NIH Grant Nos. R01DC008365 to R.Y.L. and R01DC02932 to J.R.E.]

3pSC21. Discrimination of voicing on the basis of temporal envelope cues in 6-month old infants. Josiane Bertoncini, Laurianne Cabrera, and Christian Lorenzi (Lab. Psychologie de la Percept., CNRS-Université Paris Descartes, ENS, 45 rue des Sts Pères, 75006 Paris, France, josiane.bertoncini@parisdescartes.fr)

Several studies indicate that profoundly deaf children receiving a cochlear implant (CI) under the age of 2 years are able to develop linguistic skills at a rate equal to similarly aged children with normal hearing. CI devices deliver temporal-envelope (E) cues in speech over a small number of frequency channels. This suggests that infants are able to use efficiently E speech cues at an early age. However, little work has been done to investigate the developmental time course of the ability to use E speech cues. A recent study suggests that normal-hearing children are able to use such cues at adult levels by the age of 5 years, but information is lacking for younger children. The present study assessed the ability of 6-month old infants with normal hearing to discriminate between voiced and unvoiced stop consonants (/aba/ versus /apa/) on the basis of E cues, using a head-turn preference procedure and speech tokens processed via a tone or a noise E vocoder. The spectral and temporal resolutions of vocoders were varied to determine whether or not the ability to use E speech cues is similarly constrained in infants and adults. Preliminary results indicate that this method is applicable to young infants.

3pSC22. Infants' ability to recognize speech in the presence of amplitude-modulated background noise. Rochelle S. Newman and Giovanna Morini (Dept. of Hearing and Speech Sci. and Program in Neurosci. and Cognit. Sci., Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742, rnewman@hesp.umd.edu)

Adults experience a release of masking when the amplitude level of a masker varies [Festen and Plomp (1990) and Wilson and Carhart (1969)], especially when the variation occurs in a slow, predictable manner [Gustafsson and Arlinger (1994)]. This pattern of performance has been described as "listening in the dips" of the signal and is a form of selective attention: focusing one's attention on those time periods of the signal that are likely to be most beneficial. Infants have poor selective attention abilities for different frequency regions [Bargones and Werner (1994)]; if they similarly are less able to attend to particular points in time, they should show no advantage when the amplitude level of a masker varies. We presented infants with either their own name or a foil name in the presence of white noise; the noise was either constant in amplitude or varied at a 10-Hz modulation at a depth of 12 dB. Infants showed a disadvantage when faced with amplitude-modulated background noise (that is, they showed stronger recognition of their name when the noise had a constant amplitude than a varying amplitude), suggesting that the modulation itself may have attracted attention away from the target speech. [Work supported by NSF BCS0642294.]

3pSC23. Cross-language perception of Mandarin affricates in conversational vs careful speech. Anna C. Woods and Natasha L. Warner (Dept. of Linguist., Univ. of Arizona, Douglass Bldg., Rm. 200E, Tucson, AZ 85721)

Two affricates "zh" and "j" in Beijing Mandarin are often confused by English speakers learning the language [Dow 1972]. An experiment based

on the work of Logan in 1989 training Japanese speakers to distinguish the difference between English /ɛvr/ and /l/ was modified to include training tokens taken from native Mandarin speakers' phone-conversational speech. "j" is described by Lee and Zee [2003] as laminal pre-palatal, and "zh" is described as apical post alveolar. In careful speech, the sounds are in complementary distribution, with "j" always followed by a high front vowel or a glide. A preliminary perception experiment shows whether native Mandarin speakers are able to distinguish between the pair of affricates simply by hearing the affricate alone, or if they are only able to distinguish them when hearing the vowel information as well. In conversational speech, the following high front vowel may be deleted, so natives' perception of the two affricates may differ for casual speech. English speakers with no Mandarin experience also participate in a perception task to show the effects of training with conversational speech tokens. This study elucidates how learners acquire categories despite the variability of natural speech.

3pSC24. Effect of age on clear speech intelligibility. Rajka Smiljanic (Linguist., Univ. of Texas at Austin, CAL 501, Mailcode B5100, Austin, TX 78712)

Previous work has established that elderly adult's comprehension is more detrimentally affected by noise, multi-talker babble and the absence of contextual cues compared to younger adult's comprehension. Elderly adult's abilities to understand fast speech and to detect/recognize brief non-speech and speech sound sequences (e.g., stop gaps) also decline with age. These effects seem to persist even when hearing loss is taken into account. In this study, we explored whether there are parallel effects of age on speech production, i.e., whether elderly adult's production patterns differ from those of younger adults. Specifically, we examined whether elderly and younger adult's hyper-articulation clear speech strategies provide similar intelligibility advantage for younger adult listeners. The sentence-in-noise perception results revealed that elderly adult's clear speech increased intelligibility for young adult listeners. Importantly, the overall elderly adult's speech intelligibility and the clear speech intelligibility gain were smaller compared to that provided by the speech of younger adults. Acoustic correlates of the differences in the observed clear speech gain provided by older and younger adults are explored. These results suggest that, in addition to perceptual problems, cognitive and hearing changes due to age can affect elderly adult's speech patterns and intelligibility.

3pSC25. Acoustic correlates of clear speech vowel intelligibility for elderly hearing-impaired listeners. Sarah Hargus Ferguson (Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, safergus@ku.edu)

Earlier work on vowels in clear and conversational speech produced by a single talker suggested that the acoustic cues that underlie the superior intelligibility of clear speech for older adults with hearing loss may differ from the cues that young normal-hearing listeners find beneficial. In the present study, clear and conversational vowel stimuli produced by all 41 talkers from the Ferguson Clear Speech Database were presented in a background of 12-talker babble to 40 older adults with mild-to-moderately severe sloping sensorineural hearing loss. Acoustic analyses (vowel duration, steady-state formant frequencies, and measures of dynamic formant movement) were also performed for all stimuli. Vowel intelligibility scores for each talker in each speaking style as well as the magnitude of the clear speech vowel intelligibility effect for each talker will be compared to those obtained from young listeners with normal hearing in a previous study. Regression and other statistical analyses of intelligibility and acoustic data will also be performed to determine the acoustic correlates associated with improved clear speech vowel intelligibility for each listener group. [Work supported by NIHDCD-008886.]

3pSC26. American English adults' perception of Spanish-accented conversational and clear speech. Erika S. Levy, Paula B. Garcia, Dorothy Leone, and Jessica A. Lew (Dept. of Biobehavioral Sci., Teachers College, Columbia Univ., 525 W120th St., Box 180, New York, NY 10027)

Clear speech, an intelligibility-enhancing manner of speech production, has been shown to increase the transmission of information to various listener groups when uttered in the speaker's native language. When clear speech is used in a non-native language, it is not evident whether this mode will also improve intelligibility. The present study examined the accuracy with which adult monolingual native American English (AE) listeners identified Spanish-accented vowels produced in conversational and clear speech modes. Consonant-vowel-consonant (CVC) words were presented in the carrier sentence "Touch the CVC please," containing the AE vowels /æ, ʌ, ɑ, ɛ, i, ɪ/. The sentences were produced by a native speaker of AE and by a native speaker of Spanish judged as moderately accented by the AE listeners. Five pairs of photographs (cap-cup, map-mop, nut-knot; ship-sheep, pot-pet), representing the vowel pairs /æ-ʌ/, /æ-ɑ/, /ʌ-ɑ/, /ɪ-i/, /ɑ-ɛ/ were presented visually on a touch screen in a two-alternative forced choice paradigm. Listeners had the greatest difficulty with the /ɪ-i/ and /ʌ-ɑ/ contrasts. Clear speech mode was not as effective for non-native speech as it has been shown for native speech. Implications are considered for social and clinical settings in which a mismatch in language backgrounds is present.

Session 3pSP

Signal Processing in Acoustics and Underwater Acoustics: Sparse Approximations in Signal Processing

James C. Preisig, Cochair
Woods Hole Oceanographic Inst., Dept. Applied Ocean Physics and Engineering, Bigelow 207, Woods Hole, MA 02543-1053

Ananya Sen Gupta, Cochair
Woods Hole Oceanographic Inst., Dept. Applied Ocean Physics and Engineering, Bigelow 207, Woods Hole, MA 02543-1053

Invited Papers

1:00

3pSP1. Distilled sensing for sparse recovery. Jarvis Haupt (Dept. of Elec. and Comput. Eng., Rice Univ., 6100 Main St., Houston, TX 77005), Rui Castro (Columbia Univ., New York, NY 10027), Robert Nowak (Univ. of Wisconsin-Madison, Madison, WI 53706), and Richard G. Baraniuk (Rice Univ., Houston, TX 77005)

The study of sparsity in data-rich applications has recently garnered significant attention in the signal processing, statistics, and machine learning communities. Generally speaking, sparsity quantifies the phenomenon where, in a large collection of data, only a small fraction is significant. Techniques that exploit sparsity are currently utilized in a wide variety of applications, including conventional and medical imaging, rf communications, signal intelligence, and bioinformatics, to name a few. Recently a renewed emphasis has been placed on identifying techniques that make judicious use of sensing resources, with the goal of extracting only the relevant information in high-dimensional but sparse signals. Among these techniques, the most promising approaches typically employ some form of active sampling, or sampling with feedback. Distilled sensing highlights how active sampling can result in rather surprising (and dramatic) improvements over traditional sampling methods.

1:30

3pSP2. Efficient and adaptive sparse time-frequency representations through superposition frames. Daniel Rudoy, Prabahan Basu, and Patrick J. Wolfe (Statistics & Information Sci. Lab., Harvard Univ., Oxford St., Cambridge, MA 02138, wolfe@stat.harvard.edu)

Sparsity in signal representation requires two aspects: a large dictionary of potential basis functions that can be adapted to a variety of waveforms and an efficient means of selecting these functions and reconstructing the corresponding signal representation. This talk introduces a family of time-frequency systems termed superposition frames that enable adaptivity and also, in contrast to many adaptive analysis systems, admit the property of fast overlap-add reconstruction. Several properties of this family are proved, and specific examples of adaptation criteria in the context of sparse audio processing are discussed.

Contributed Papers

2:00

3pSP3. Random encoding for robust recovery and channel identification. William E. Mantzel (801 Myrtle St., Apt. 1, Atlanta, GA 30308, willem@gatech.edu), M. Salman Asif, and Justin K. Romberg (801 Myrtle St., Apt. 1, Atlanta, GA 30308)

By encoding a signal in a randomly chosen subspace before transmission, this transmitted signal proves robust to sparse errors in the channel. In this talk, we will further demonstrate the robustness of such signals to unknown convolutive channels. In particular, we show that if the channel is not especially resonant at any given frequency (e.g., the channel is sparse or random), we are able to simultaneously recover both the transmitted signal and the channel characteristics by solving a rank minimization in X, the rank-1 outer product of the two unknown vectors, subject to the receiver observations. We demonstrate favorable simulated performance for sparse and random channels. Using a parallel approach to classical compressed sensing theory, we prove a restricted isometry property about our linear operator that guarantees a high probability of recovery when the number of observations is at least linear with the dimension of the message and the length of the channel.

2:15

3pSP4. Shallow water channel tracking using gradient-based mixed norm optimization techniques. Ananya Sen Gupta and James Preisig (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The underwater acoustic channel is challenging to estimate and track due to channel dynamism, the ill-conditioned nature of the estimation problem, as well as the need to optimize over a complex field. An extension of previous work is presented where the sparse channel coefficients in the delay and delay-Doppler spread domains are estimated using mixed norm convex optimization. The estimation algorithm is extended to track the time-varying channel using a gradient descent approach, trading off optimality for tracking efficiency. A computationally inexpensive approximation of the dynamic gradient is used rather than re-computing the precise gradient every time. Results presented are based on field data collected over 15-m depth and a range of wave heights and distances using the normalized prediction error as the performance metric. Comparison of performance between tracking the delay-only and delay-Doppler coefficients is presented for different weightings on the mixed norm between L1 and L2 measures and over different lags. Performance and computational efficiency of the proposed

tracking algorithm are compared to conventional sparse estimation techniques. [Work supported in part by ONR Grants N00014-05-10085 and N00014-07-10184 and in part by Postdoctoral Scholar Program at the Woods Hole Oceanographic Institution, with funding provided by the Doherty Foundation.]

2:30

3pSP5. Convex hull method of acoustic field interpolation. Brian R. La Cour (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Empirical knowledge of the local acoustic field typically comes in the form of time samples at discrete points in space corresponding to the locations of, say, omnidirectional microphones or hydrophones on an underwater array. To determine the acoustic field at an arbitrary point in space requires either propagation modeling, signal interpolation, or some combination of the two. In this paper, a method of field interpolation is presented which estimates the acoustic field at any point within the convex hull of all sample points. This method is compared against more traditional inverse beamforming methods using both simulated and measured data.

2:45

3pSP6. Compressive sensing in underwater acoustics. Geoffrey F. Edelmann and Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, geoffrey.edelmann@nrl.navyl.mil)

The performance of compressive sensing for determining frequency tonals is shown for several cases. The sampling requirements for infinite signal-to-noise ratio (SNR) is shown first. A detection threshold is developed for the case of additive noise with known noise level. The number of samples needed per tonal is shown as a function of SNR. The effects of temporal variability in the form of fading due to motion through an oceanic waveguide is discussed. [Work supported by the Office of Naval Research.]

3:00

3pSP7. Compressive matched field processing. William Maetzel, Justin Romberg (School of Elec. and Comput. Eng., Georgia Inst. of Technol., 777 Atlantic Dr. NW, Atlanta, GA 30332-0250), Karim G. Sabra (Georgia Inst. of Technol., Atlanta, GA 30332-0405), and William Kuperman (UC. San Diego, La Jolla, CA)

Sound source localization in shallow water environment is commonly done using matched field processing (MFP). MFP is usually implemented by systematically placing a test point source at each point of a search grid, computing the acoustic field (replicas) at all the elements of the array and then correlating this modeled field with the data from the real point source whose localization is unknown to determine the best-fit location. However, a direct implementation of MFP (i.e., brute force search) over a large grid space—or search area—is computationally demanding especially in the presence of complex propagation environments. We formulated instead the localization problem of a few acoustic sources as the sparse approximation of the measured signals in a specific dictionary of atoms obtained from discretization of this sparse search space. Spatial random projections are performed to efficiently span the search space. This compressive MFP approach allows for significant computation time savings with a computational cost increasing with the number of random projections instead of the number of search grid points. The performance of this approach will be illustrated using numerical and experimental data in a shallow water waveguide.

ASA/INCE Plenary Session and Awards Ceremony

Whitlow W. L. Au, Cochair
President, Acoustical Society of America

James K. Thompson, Cochair
President, Institute of Noise Control Engineering

Presentation of INCE-USA Awards

INCE Student Paper Prizes

Presentation of Certificates to New Fellows of the Acoustical Society of America

Frank H. Brittain
Jean-Yves Chapelon
Egbert de Boer
Richard L. Freyman
Paul M. Gammell
Karl Grosh
Allard Jongman
David P. Knobles
Pascal P. Laugier
Ruth Y. Litovsky
Oleg I. Lobkis
Enrique Lopez-Poveda
Paul E. Nachtigall
Susan N. Nittrouer
Brian Roberts
Mark Sheplak
Barbara Shinn-Cunningham
Mitchell S. Sommers
Peter J. Stein
Keith A. Wear

Presentation of Acoustical Society Awards

ASA Certificates of Appreciation to Trustees of the Acoustical Society Foundation

Leo L. Beranek
Mahlon D. Burkhard
Richard H. Campbell
Stanley L. Ehrlich
Paul B. Ostergaard
Richard Stern
Murray Strasberg
Louis C. Sutherland
Robert D. Frisina
Katherine S. Harris
Juliette W. Ioup
William W. Lang
Jiri Tichy
Gregory C. Tocci
George P. Wilson

R. Bruce Lindsay Award to Kent L. Gee

von Békésy Medal to William S. Rhode

Helmholtz-Rayleigh Interdisciplinary Silver Medal to Ronald A. Roy

Gold Medal to Jiri Tichy

Session 3eED

Education in Acoustics and ASA Women in Acoustics Committee: Listen Up and Get Involved

Marcia J. Isakson, Chair
Univ. of Texas at Austin, Applied Research Labs., 10000 Burnet Rd., Austin, TX 78713

Chair's Introduction—5:30

Invited Papers

5:35

3eED1. Hands-on demonstrations for Project Listen Up: Education outreach. Hannah E. Allaire, Bethany L. Carlson, Julie Devlin, Alaina D. Dibiasie, Krysta L. Porterlott, Erica B. Renzhofer, Janet S. Voneiff, Jessie Carman, and Murray S. Korman (Dept. Phys., U.S. Naval Acad., Annapolis, MD 21402)

Midshipmen from the General Science Major will be getting involved in an ASA education outreach effort by presenting a number of acoustical demonstrations geared to promote a hands-on learning experience for middle- and high-school age girl scouts. The demos are designed to visualize certain sound wave effects that will be explained by the Midshipmen (role models) who hands-off the controls of the apparatus to the students who will be free to participate and make their own scientific discoveries. The demonstrations will be (1) sympathetic vibrations of matched xylophone bars or matched tuning forks mounted in quarter-wave resonant open-ended boxes, (2) tuning fork excitation of standing waves in a sound resonance tube with a variable water column, (3) small inexpensive speaker placed in a hole in a baffle or inside a megaphone generating enhanced sound by eliminating destructive interference effects, (4) interference from two separated loud speaker sources measured vs microphone location, and transverse wave visualizations using a (5) torsion wave machine and (6) a small ripple tank. The traveling and standing wave patterns in (5) and the reflection, refraction, interference, and scattering effects in (6) are helpful wave visualizations and aid in the understanding of longitudinal sound wave effects.

5:41

3eED2. How musical instruments make sounds: An exploration of standing waves and resonance frequencies. Tracianne B. Neilsen (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602, tbn@byu.edu)

A basic understanding of the production of sound waves can be obtained by studying standing waves and resonance frequencies. A vibrating string will settle into a steady, standing wave pattern when it is driven at one of its resonance frequencies. The resonance frequencies depend on the string length and the tension applied to the string. Similarly the resonance frequencies of tubes are determined by the tube length and whether the ends are open or closed. In a series of hands-on demos, we will explore the factors that influence the creation of standing waves by exciting the resonances of strings, tubes, rods, a metal plate, slinky, and a wine glass. These simple models provide insight into how musical instruments produce sound.

5:47

3eED3. Chlandi pattern demonstration. Marcia J. Isakson (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

Chlandi patterns will be demonstrated using sand on a plate and an air acoustic source. As the source frequency is changed, different nodal patterns will become evident on the plate. In this manner, the different vibrational modes of the plate can be qualitatively determined.

5:53

3eED4. The singing shoebox: A $5 loudspeaker project. Scott P. Porter, Daniel J. Domme (The Graduate Program in Acoust., The Penn State Univ., 201 Appl. Sci. Bldg., State College, PA 16802), and Jeffrey S. Whalen (The Penn State Univ., State College, PA 16802)

Moving-coil loudspeakers typify the interdisciplinary nature of acoustics. In order to reproduce sound, these devices employ principles of electricity, magnetism, mechanics, and acoustics. The widespread use of loudspeakers has made them familiar to students and thus a valuable opportunity to introduce students to acoustics. In this paper, the authors demonstrate a loudspeaker-enclosure system that is easily built from scratch. The system is mostly constructed from common household supplies, making it low-cost and accessible to a wide audience. In addition, this project is well-suited for use by the K-12 educator as the content can be scaled to fit a variety of different academic levels. To this end, the speaker's design and construction will be presented and its relevance as an educational demonstration discussed. Finally, the loudspeaker will be auditioned and its performance demonstrated.

5:59

3eED5. Longitudinal wave demonstrations. Michelle E. Swearingen (Construction Eng. Res. Lab., USA ERDC, P.O. Box 9005, Champaign, IL 61826, michelle.e.swearingen@usace.army.mil)

In a wave, something always moves back and forth. A longitudinal wave is one that travels along the direction of propagation. In this demonstration, two different devices will be used to illustrate longitudinal propagation. An aluminum bar is used to produce transverse waves. A slinky will be used to visually show longitudinal propagation along the slinky. Finally, the aluminum bar is again used to show differences in longitudinal and transverse wave propagations.

6:05

3eED6. Demonstration of an air column resonance with a corrugated tube. Michelle C. Vigeant (Dept. Mech. Eng., Acoust. Prog. and Lab., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

Standing waves, or resonances, will occur in an open-ended tube when the air in the tube is excited at a particular frequency or its harmonics. The resonant frequencies are a function of the tube length. The first standing wave, fundamental mode or first harmonic, is set-up with antinodes at each end of the tube, as the air is free to move at the ends, with one node in the middle of the tube. Each successive standing wave has an additional anti-node and node, with the second harmonic containing three, antinodes and two nodes, the third containing four and three, respectively, and so on. This phenomenon of an air column resonance will be demonstrated using corrugated tubes. By spinning the tube overhead at a particular speed, the resonance frequencies will be excited. For slower speeds, lower frequencies will be heard, while the pitch will increase for faster speeds. Instrumental examples of open-ended tubes include the flute and organ pipes.

6:11

3eED7. Amplification. Ananya Sen Gupta (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

In this experiment the students will see a demonstration of how sound waves can be amplified. The setup of the experiment is simple: a tuning fork, a table, and a music box. First the tuning fork is struck by itself and the level of the sound emitted is noted. Then the tuning fork is struck again and held against a table. The amplification of the sound due to propagation of acoustic pressure waves through the larger surface area of the table is noted. The students are encouraged to participate and give their own reasoning for sound amplification. If time permits, students will experiment with the music box as well.

6:17

3eED8. Experiment on beats. Ananya Sen Gupta (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

When two sounds of nearly the same frequency get together they will alternately add up or cancel out. If they start out together, after a while they will no longer be together, and the peak of the one will meet a trough of the other and they will cancel. The resulting fluctuations are known as beats. The beat frequency is just the difference of the two frequencies. The students will observe and experiment with beats generated by sounding an 880 tuning fork at the same time as an 883 tuning fork.

6:23

3eED9. The tin-can telephone: An example of sound propagation and communication for Project Listen Up. Jason D. Sagers, Andrew R. McNeese, and Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

The tin-can telephone can be used to illustrate the basic concepts of sound waves and sound-structure interaction. It can also be used to illustrate advanced concepts of frequency response and speech intelligibility in acoustic communications. For the proposed demonstration, students will construct a tin-can telephone from household materials and observe its performance from a variety of acoustic tests. Students will observe sound transmission and discover that some sounds are transmitted with greater clarity than others. A measured frequency response function for several tin-can telephones constructed from different materials will be provided to the student to illustrate the concept of frequency response and speech intelligibility.

6:29

3eED10. Ultrasonic levitation. Cecille Pemberton Labuda (Univ. of MS/NCPA, 1 Coliseum Dr., Univ. MS 38677, cpembert@olemiss.edu)

Ultrasonic standing waves are set up between a horn and a brass plate. Because the wavelength is short, several wavelengths fit in the space between the horn and the plate. There are locations in the standing wave pattern where the forces associated with the sound pressure and the motion of the air molecules are opposite in direction to the gravitational force. A light ball placed at any of these locations will levitate if the mass of the ball is such that the gravitational force on the ball is equal and opposite to the forces associated with the pressure and the motion of the air molecules.

6:35

3eED11. Fourier synthesis. Kyoko Nagao (Ctr. for Pediatric Auditory and Speech Sci., Nemours A.I. duPont Hospital for Children, 1600 Rockland Rd., Wilmington, DE 19803, nagao@asel.udel.edu)

When two or more simple sound waves that differ in frequency and amplitude are added together, a complex wave is generated. This is called Fourier synthesis, named after a French mathematician Joseph Fourier. Many sounds we hear every day are complex waves. For example, the sound wave produced by the vowel "ah" is a complex wave. By combining simple waves with appropriate values, we can generate speech with a computer. In this experiment, you will synthesize a complex sound wave by adding two or more pure tones using a computer program. The ultimate goal here is to create a complex tone that sounds like the vowel ah. You will learn important acoustic components used in speech sounds.

6:41

3eED12. Acoustic demonstrations for the iPod generation. Shawn F. Johnson and Juan I. Arvelo (Appl. Phys. Lab., The Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

Today's "iPod generation" can gain a unique and interactive exposure to physics by leveraging their musical interests, and their iPod, to explore the science of sound. Inside certain models reside a mini-computer and a suite of acoustic and vibration sensors, including loudspeakers, microphones, and accelerometers. Combined with free or inexpensive software applications or "apps," these highly portable devices can offer a tangible means to help students understand many physics, audio, and acoustics concepts that they can explore both inside and outside the classroom. We will demonstrate several apps that are currently available, for example, a seismometer, signal generator, noise level meter, spectrum analyzer, oscilloscope, and various musical instruments. Demonstrations utilizing an iPod Touch are guaranteed to excite students of all ages.

6:47

3eED13. Doppler effect. Megan S. Ballard (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin TX 78758, meganb@arlut.utexas.edu)

The Doppler effect describes the change in frequency observed for a moving source and/or receiver. A common example of the Doppler effect involves a moving vehicle sounding a siren or horn as it approaches, passes, and recedes from an observer. The observed frequency (compared to the emitted frequency) is higher during the approach, is identical at the instant of passing by, and is lower during the recession. In this presentation, the Doppler Effect is demonstrated using a buzzer embedded in a Nerf ball. A difference in pitch can be heard when the ball is thrown to or from an individual. A second demonstration involves twirling a tuning fork from an attached string. An increase/decrease in frequency can be heard as the tuning fork moves toward/away from an observer.

6:53

3eED14. Phonation: Mickey Mouse and Darth Vader. Tessa Bent (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, tbent@indiana.edu)

During voicing, the vocal folds are brought together and then caused to vibrate by air being pushed out of the lungs. In order to illustrate several phenomena regarding phonation, three demonstrations will be provided. First, an experiment with balloons will demonstrate how people can change the pitch of their voice by stretching their vocal folds. Second, the aerodynamic forces involved in phonation will be demonstrated with the use of soda cans and a straw. Last, high-speed videos of the vocal folds in action will be shown.

6:59

3eED15. Listen up and get involved: Standing waves—The unbalanced motor and musical string. Pamela J. Harght (4006 Speedway, Austin, TX 78751, pam@baiaustin.com)

The experiments that will be demonstrated are Standing Waves—The Unbalanced Motor and Standing Waves—Musical String. The unbalanced motor experiment involves a small electric motor that provides tension in the string. Students will observe the effect of holding the string and slowly sliding their hand toward the unbalanced motor. For the musical string experiment, students will observe musical intervals (do-me, do-sol, do-do, etc.) with a tensioned piano string and the science behind musical notes.

7:05

3eED16. Fourier synthesis. Helen M. Hanson (Dept. ECE, Union College, 807 Union St., Schenectady, NY 12308, helen.hanson@alum.mit.edu)

Fourier synthesis is the act of building a complex waveform from simple sine waves. Fourier analysis, on the other hand, decomposes a complex waveform into simple sine waves. Both Fourier synthesis and analysis are powerful tools in many scientific and engineering fields, allowing one to understand waveforms of all types in terms of their frequency content. In this experiment, Fourier synthesis will be brought to life as an acoustic waveform is built by combining several harmonic waves. A keyboard will generate the waves, which will then be observed using a computer.

7:11

3eED17. The energy flow for a spherical acoustic lens: Experimental results avoiding interference effects. Kendez C. Parker and Cleon E. Dean (Dept. Phys., Georgia Southern Univ., P.O.B. 8031, Statesboro, GA 30460-8031, cdean@georgiasouthern.edu)

A simple classroom demonstration consists of a weather balloon filled with carbon dioxide, a sound source, and a microphone. Since the speed of sound is slower in carbon dioxide than in air at room temperature and pressure, the balloon acts as a positive spherical acoustic lens. Preliminary experimental results have been presented previously [C. E. Dean and J. P. Braselton, "The energy flow for a spherical acoustic lens: ray and wave methods vs. experiment.," J. Acoust. Soc. Am. **125**, 2627 (2009)]. The possibility of interference effects from the reflection of sound off surfaces was brought up in the ensuing discussion. The current results have been measured in a way that minimizes the effect of interference due to reflections off walls, floor, or other surfaces.

7:17

3eED18. Standing waves on a vibrating string. Dorea R. Ruggles (Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215)

Most sounds are made up of many waves added together. These waves, called harmonics, are all related to each other because their frequencies are integer multiples of the lowest frequency, the fundamental. Adding together different amounts of each harmonic is part of what makes different sounds unique. The individual harmonics of a vibrating string are easy to see, one at a time, by creating a standing wave on the string and adjusting the driving frequency. The frequency where the fundamental, one-looped standing wave appears on the string depends on the string's length, mass, and tension, but if these factors are all kept constant while the driving frequency is increased, a new loop will be added to the standing wave for each multiple of the original frequency. This is an exciting way to see waves, observe their properties, and really understand how frequency, wavelength, and other properties are all interrelated.

7:23

3eED19. Ocean acoustic tomography: Live from the Philippine Sea. Lora J. Van Uffelen (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., Mail Code 0225, La Jolla, CA 92093-0225) and Kathleen E. Wage (George Mason Univ., Fairfax, VA 22030)

In this live interactive video session, scientists onboard the R/V Roger Revelle in the Philippine Sea off the coast of Taiwan describe the basic concept of sound propagation in the ocean in the context of an ongoing acoustic mooring deployment. Six acoustic source moorings and a distributed hydrophone array will be deployed during the course of the month-long cruise. The scientists will give an introduction to how research is carried out at sea and will describe basic concepts and advantages of using acoustics as a tool to learn about the ocean environment. They will show examples of oceanographic instrumentation, including conductivity-temperature-depth sensors, acoustic sources, and hydrophone receivers, and will describe the process of deploying acoustic moorings and acquiring oceanographic data. The girl scouts will learn how sound is used to map the 6-km deep ocean floor as well as how tomographic moorings can be used to sense the ocean interior. There will be the opportunity to interact with scientists on the research ship and ask questions about ocean acoustics or life at sea. [Field work supported by ONR.]

7:29

3eED20. Standing waves on a spring. Holly A. Smith and Tara L. Tubbs (Grad. Prog. in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802)

This demonstration shows how standing waves on a spring can be generated by having two people simulate boundary conditions. If one person holds the end of the spring at rest while the other drives the opposite end in an up and down motion, the spring will begin to oscillate. By having the person driving the spring adjust her rate of motion, she will eventually find a rate of motion that produces a single standing wave on the spring. This rate is the fundamental frequency, and it is dependent upon the tension in the spring as well as the spring's mass per unit length. The fundamental frequency can be measured by timing the up and down motion of the standing wave for 20 cycles and dividing 20 by the recorded time, giving you the fundamental frequency in cycles per second. Other harmonics of the spring can be produced by increasing the driving rate to integer multiples of the fundamental frequency.

State of the Art of Supersonics Aircraft Technology–What Has Progressed in Science Since 1973

Victor W. Sparrow, Cochair
Pennsylvania State Univ., Graduate Program in Acoustics, 201 Applied Science Bldg., University Park, PA 16802

Laurette Fisher, Cochair
Federal Aviation Administration, 800 Independence Ave., SW, Washington, D.C. 20591

This panel discussion is an FAA public meeting on the advances in supersonic aircraft technology and related research in the noise generated during supersonic flight. The purpose of this session is to raise public awareness of the continuing technological advances in supersonic aircraft technology, and for the FAA, the National Aeronautics and Space Administration (NASA), and industry to get feedback from interested persons. The session will include presentations on current research programs and a question and answer session. Public involvement is essential in any future definition of an acceptable new standard that would allow supersonic flights over land. The discussion is open to all interested parties, and there are no fees to attend this session. [The FAA thanks the ASA for the opportunity to include this panel discussion as a part of the ASA's 159th meeting, joint with Noise-Con 2010.]

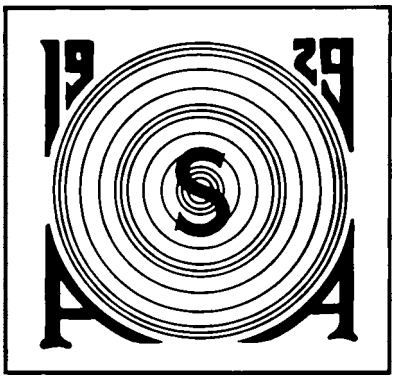
OPEN MEETINGS OF ASA TECHNICAL COMMITTEES

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held immediately after the Social Hours. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings, including proposals for special sessions, workshops, and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions. The Technical Committee on Signal Processing in Acoustics will meet tonight in Grand Ballroom III/IV beginning at 7:30 p.m.

ACOUSTICAL SOCIETY OF AMERICA

R. BRUCE LINDSAY AWARD



Kent L. Gee

2010

The R. Bruce Lindsay Award (formerly the Biennial Award) is presented in the Spring to a member of the Society who is under 35 years of age on 1 January of the year of the Award and who, during a period of two or more years immediately preceding the award, has been active in the affairs of the Society and has contributed substantially, through published papers, to the advancement of theoretical or applied acoustics, or both. The award was presented biennially until 1986. It is now an annual award.

PREVIOUS RECIPIENTS

Richard H. Bolt	1942	Gilles A. Daigle	1988
Leo L. Beranek	1944	Mark F. Hamilton	1989
Vincent Salmon	1946	Thomas J. Hofler	1990
Isadore Rudnick	1948	Yves H. Berthelot	1991
J. C. R. Licklider	1950	Joseph M. Cuschieri	1991
Osman K. Mawardi	1952	Anthony A. Atchley	1992
Uno Ingard	1954	Michael D. Collins	1993
Ernest Yeager	1956	Robert P. Carlyon	1994
Ira J. Hirsh	1956	Beverly A. Wright	1995
Bruce P. Bogert	1958	Victor W. Sparrow	1996
Ira Dyer	1960	D. Keith Wilson	1997
Alan Powell	1962	Robert L. Clark	1998
Tony F. W. Embleton	1964	Paul E. Barbone	1999
David M. Green	1966	Robin O. Cleveland	2000
Emmanuel P. Papadakis	1968	Andrew J. Oxenham	2001
Logan E. Hargrove	1970	James J. Finneran	2002
Robert D. Finch	1972	Thomas J. Royston	2002
Lawrence R. Rabiner	1974	Dani Byrd	2003
Robert E. Apfel	1976	Michael R. Bailey	2004
Henry E. Bass	1978	Lily M. Wang	2005
Peter H. Rogers	1980	Purnima Ratilal	2006
Ralph N. Baer	1982	Dorian S. Houser	2007
Peter N. Mikhalevsky	1984	Tyrone M. Porter	2008
William E. Cooper	1986	Kelly J. Benoit-Bird	2009
Ilene J. Busch-Vishniac	1987		

CITATION FOR KENT L. GEE

. . . for contributions to the fields of jet noise propagation, nonlinear acoustics, and active control of fan noise

BALTIMORE, MARYLAND • 21 APRIL 2010

Kent Leonard Gee was born in Cleveland, Ohio and grew up in nearby Bay Village. A career in education seemed likely for him, as his father was a faculty member at Cleveland State University, and his mother was an elementary school teacher. In high school, Kent studied physics for two years, taught by a great instructor, and decided that he wanted to be a teacher. In 1994, Kent moved to Provo, Utah to attend Brigham Young University (BYU) on a full-tuition scholarship. After one year of schooling he took a two-year break to serve in southern Spain as a missionary for his church. He then returned to BYU to further pursue his education.

Kent is a well-rounded individual. He was blessed with speed and ran track for BYU for two years until injuries led him to give up that activity. Kent enjoyed playing the trumpet and piano, even composing some music occasionally. Most importantly for his future career, it was during this time that he discovered the field of acoustics. He realized he had a passion for the subject, and began taking graduate acoustics courses at BYU during his junior year in 1999-2000.

Armed with his new-found training in acoustics, Kent headed back home for the summer of 2000 for an internship at the Acoustical Testing Laboratory at NASA Glenn Research Center under the direction of Beth Cooper. Although he primarily worked on instrumentation and identifying flight-approved acoustic materials for the International Space Station, it was during also this internship that Kent was first exposed to the field of nonlinear acoustics.

Kent completed his BS in Applied Physics with an acoustics emphasis in 2001, graduating magna cum laude with minors in mathematics and Spanish. He immediately continued at BYU in their graduate program, working with Scott Sommerfeldt on developing a new approach for active noise control of small axial fans. Kent developed a novel compact control system that locates all acoustic control sources and sensors in the near field—very challenging for achieving global active control. This work was later published in the *Journal of the Acoustical Society of America* (*JASA*) and in the *Noise Control Engineering Journal* (*NCEJ*). The *NCEJ* paper was recognized in 2004 by the Institute of Noise Control Engineering (INCE-USA) with its biennial Martin Hirschorn IAC Best Paper award. Kent also developed his multitasking skills at BYU. In addition to his thesis research on axial fans, Kent took on the lead role in developing a three-dimensional intensity scanning system for NASA Glenn, which developed out of his earlier internship.

After graduating from BYU with his M.S. in Physics in 2002, Kent moved to Penn State to pursue a Ph.D. in the Graduate Program in Acoustics. Vic Sparrow was able to attract him with a University Graduate Fellowship in the Fall of 2002 and a Graduate Research Assistantship related to the propagation of high-amplitude jet noise beginning in the summer of 2003. During the joint Penn State/Wyle Laboratories project funded by the U.S. Department of Defense (DoD), Kent was able to rub shoulders with experts both in academia and industry. Penn State faculty members exposed him to different perspectives on jet noise and acoustical measurement techniques. Wyle staff members also provided sage advice and assistance, as well as the opportunity to participate in propagation measurements on an F-22 Raptor aircraft at Edwards Air Force Base to provide benchmark experimental data.

Kent's Ph.D. research focused on accurately predicting the far-field propagation of high-amplitude military jet noise—something not accomplished previous to Kent's work. This is an extremely important topic due to the rapid encroachment of populated areas around military air bases. Extending an approach originated by David Blackstock and his students, Kent was the first person to make accurate predictions based on the physics of the problem, rather than empirical guesses. The propagation of nonlinear jet noise is a challenging topic, but Kent makes it look easy. Multiple peer-reviewed publications have

come out of the work and his methods have been used directly by Wyle to produce a new DoD military jet noise prediction model. Certainly through this research, Kent has become a superlative young contributor to acoustics.

After graduating with his Ph.D. in the summer of 2005, Kent had several options to consider, but decided to return to BYU in a faculty position—first as a Visiting Assistant Professor and then in a permanent position since January 2006. He has been an important contributor to building the acoustics program at BYU, which currently has about 30 students enrolled—about half graduate and half undergraduate students. Kent has supervised four M.S. students, and is currently supervising an additional two M.S. students and one Ph.D. student. Perhaps equally impressive for the future of acoustics, he has supervised 16 undergraduates in acoustics research, at least half of whom have gone on to pursue graduate degrees in acoustics.

Kent's multitasking skills are evidenced in his research interests, where he has worked on studying high-amplitude jet noise and rocket noise, near-field acoustical holography, Gatling gun noise, active noise control, and acoustics demonstrations. He has published much of this work, with close to half of his papers appearing in *JASA*. Outreach in acoustics is important to Kent. He has given tours of BYU's acoustics facilities to hundreds of K-12 students and participates in the American Association of Physics Teachers to help educators understand how acoustics can be a wonderful mechanism to teach physics principles.

Kent has participated regularly in activities of the ASA since 2002. He has presented his work at the majority of the meetings since then, is a member of the Physical Acoustics and Noise technical committees, and has been particularly active in the Committee on Education in Acoustics. He received two student paper awards during his student years and the Noise Young Presenter award shortly after graduating. His students now regularly receive paper awards. Kent also serves as the Associate Editor for Noise for ASA's *Proceedings of Meetings on Acoustics*.

Kent and his wife, Alicia, are the parents of five children—one born while Kent was a student at BYU, two born while he was studying at Penn State, and two born since he returned to BYU. Again multitasking comes into play, as Kent balances his time between his family, which is very important to him, and the passion he feels for his chosen discipline.

VICTOR W. SPARROW
SCOTT D. SOMMERFELDT

ACOUSTICAL SOCIETY OF AMERICA

VON BÉKÉSY MEDAL

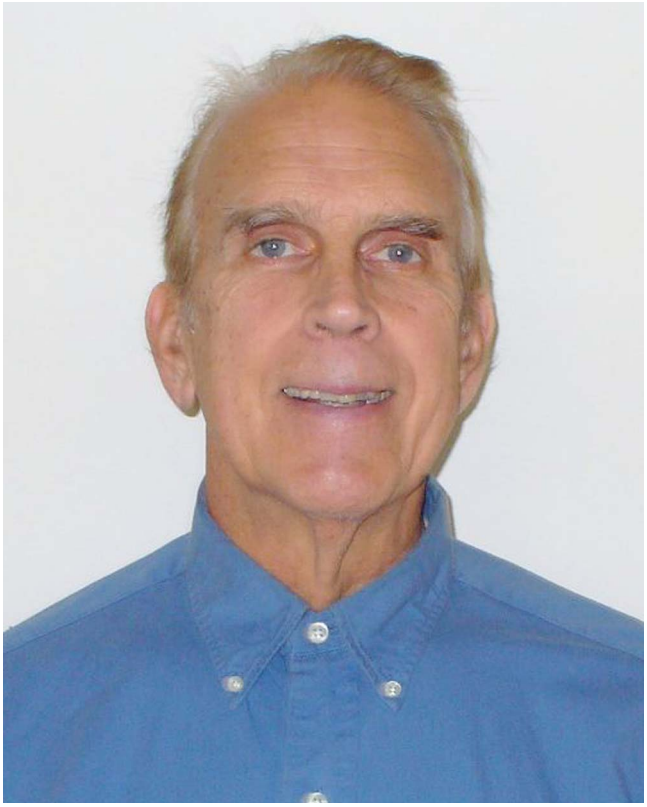
William S. Rhode

2010

The von Békésy Medal is presented to individuals, irrespective of nationality, age, or society affiliation, who have made outstanding contributions to the area of psychological or physiological acoustics as evidenced by publication of research results in professional journals or by other accomplishments in the fields.

PREVIOUS RECIPIENTS

Jozef J. Zwislocki	1985
Peter Dallos	1995
Murray B. Sachs	1998



CITATION FOR WILLIAM S. RHODE

. . . for discovering nonlinear basilar-membrane responses and for contributions to cochlear-nucleus functional circuitry

BALTIMORE, MARYLAND • 21 APRIL 2010

By 1960 the basic "tonotopic" pattern of vibrations set up in the mammalian cochlea by sound was understood, thanks to the life's work of Nobel-Laureate Georg von Békésy. His research showed that a pure tone produces a mechanical vibration on the cochlea's basilar membrane, at that sound's frequency, which propagates along that membrane from the cochlea's input port at the base toward its apex. The lower the tone's frequency, the further along the basilar membrane that traveling wave progresses until it hits an apparent "resonance," crests, and vanishes. Yet puzzles remained. These vibrations were uniformly linear over most of the audible intensity range, and they were tiny: the estimated peaks at hearing thresholds were on the order of 10^{-3} nanometers. Could such sub-atomic motion possibly stimulate receptors? Moreover, von Békésy's measurements, made with ordinary (wideband) light, showed only very shallow (low-Q) resonances on the basilar membrane. How then account for the exquisite frequency sensitivities displayed by mammalian auditory-nerve neurons and, psychoacoustically, by mammalian listeners?

Little further progress was made until 1967, when a brief paper in *Science* by Johnstone and Boyle of Australia introduced a more sensitive technique for measuring vibrations of the basilar membrane using a quantum-mechanical (Mössbauer) effect. William (Bill) Rhode, looking for a dissertation project at that time, decided to follow up that initial finding. This necessitated combining very exacting surgical techniques with mastery of unfamiliar gamma-ray equipment, while utilizing and programming one of the very first dedicated laboratory computers (a LINC). Displaying most extraordinary multidisciplinary skill and tenacity over scores of experiments, Bill succeeded in making reliable measurements of basilar-membrane vibrations in the basal (high-frequency) region to pure tones using the new technique. His initial results, published in the *Journal of the Acoustical Society of America* (*JASA*) in 1971, had a stormy reception. They showed, contrary to von Békésy's results, that the traveling-wave vibrations on the basilar membrane were markedly nonlinear, having peaks that were more sharply tuned at low sound intensities than at higher ones. It took almost 10 years for Bill's results to be generally accepted.

Since then, work in several laboratories, including his own, has confirmed that vibrations in the living and undamaged mammalian basilar membrane are indeed strongly nonlinear and sharply tuned at low intensities. In essence, each segment of the cochlear partition contains a "cochlear amplifier" that operates around its "resonant" frequency with amplifications at low intensities of up to 60 dB (compared to a post-mortem cochlea). Bill's pioneering role in discovering and exploring this "amplification" is well recognized. His 1971 paper alone has been referenced at least 625 times. More formally, he received the Samuel Talbot Award of the Institute of Electrical and Electronic Engineers for the best biomedical engineering Ph.D. thesis (1970), and in 2001 Bill received the Award of Merit from the Association for Research in Otolaryngology.

Parallel with his continuing work on cochlear mechanics, in the late 1970's Bill tackled the equally daunting task of understanding how neurons in the mammalian cochlear nucleus (CN), the first integrative stage of the auditory pathway, handle the information they receive from the auditory nerve. At the time, there existed substantial work on the anatomy of the cochlear nucleus. Anatomists had shown that it contains more than a dozen different types of neurons of which many are intermingled. Over the previous decade auditory electrophysiologists, too, had explored the cochlear nucleus, finding that the responses of its neurons to sounds were complex and quite varied. However, the fact that CN neurons with differing response patterns are not spatially segregated made it impossible in most cases to definitively associate particular response types with anatomical types in extra-cellular recordings. Bill reasoned that establishing such associations would require making intra-cellular recordings of responses to sounds of individual neurons and then labeling them so that the responses of individual neurons could be unambiguously assigned to particular anatomical types. Once again he tackled a fundamental

problem with a challenging technique: it is very difficult indeed to impale small neurons in animals whose every heartbeat and every breath threatens to pull the cell away from the electrode. With great experimental finesse, Bill and two junior colleagues succeeded in correlating anatomical cell types with the various response types to sound. Their pioneering work on both the dorsal and ventral divisions of the cochlear nucleus opened the door to major advances in our understanding of how the different sub-circuits in these nuclei process auditory information.

In addition to the twin peaks of his career, Bill has kept up a steady stream of publications on how auditory-nerve neurons and CN neurons respond to a wide variety of sound stimuli. He has been particularly interested in finding out to what extent the vibrations of the cochlear partition can account for the responses of these neurons. In the same vein, he has just recently published a paper showing correlations between (nonlinear) distortion-product otoacoustic emissions and related components in basilar-membrane vibrations.

Bill received his Ph.D. in Electrical Engineering from the University of Wisconsin-Madison (UW) in 1969, and has worked there ever since. During his tenure in Madison, where he rose to the rank of Professor of Physiology (now Emeritus), Bill was also a key player in the University's development of laboratory computing. He was first the Assistant Director (1970-72), then Director (1972-78), of the Laboratory Computer Facility, which introduced laboratory computing to the UW Medical School. He also was responsible for one of the first digital stimulus systems, which he built with the aid of the Medical Electronics shop and sent to many colleagues across the country.

In summary, Bill Rhode has made extensive and seminal contributions to the field of auditory physiology. Remarkably, he has done much of this work with his own hands, usually working, throughout his career, either by himself or with just one or two junior colleagues, most of whom have gone on to productive careers of their own in auditory physiology. For the truly stellar quality of his work, Bill's many friends join with me in congratulating him.

C. DANIEL GEISLER

ACOUSTICAL SOCIETY OF AMERICA

HELMHOLTZ-RAYLEIGH INTERDISCIPLINARY

SILVER MEDAL
in

Biomedical Ultrasound/Bioresponse to Vibration
and
Physical Acoustics

Ronald A. Roy

2010

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

CITATION FOR RONALD A. ROY

. . . for contributions to the fields of biomedical ultrasound and nonlinear bubble dynamics

BALTIMORE, MARYLAND • 21 APRIL 2010

Ronald Aurele Roy grew up in Lewiston, Maine, a former mill town on the banks of the Androscoggin River, in a Catholic, Franco-American family. French was spoken at home and continued throughout the family's adult years. Ron's father worked in a shoe shop and his mother in a textile mill. Even though money was tight, Ron's mother bought him science kits, a microscope, helped him set up science experiments in his room, and accompanied him to science fairs all over Maine. It is apparent that Ron has been an avowed scientist since he was a child.

Ron attended St. Dominic Regional High School: Angela Mickalide, a classmate, recalls, "Even as a high school student Ron was brilliant. He grasped complex ideas quickly, and posed innumerable questions that his teachers could rarely answer. However, his Achilles' heel was his atrocious spelling [which continues to this day]. Ron would tutor me daily in physics. On tests, I often earned a hundred percent, but Ron only received ninety-nine percent because he would spell words like 'parallel' incorrectly. As a result, I was awarded the Physics Prize at the Senior Award Ceremony. But don't feel too badly for him—Ron earned the Mathematics Prize!"

After high school, Ron worked at General Electric for a year before attending the University of Maine (UM) in Orono. After his junior year, Larry Crum received a phone call from a friend at UM, Orono who asked whether Larry had any openings for a first-rate undergraduate physics student. Within a few days, Ron showed up at the University of Mississippi and began working in Larry's lab, starting a partnership that has continued to this day. Ron soon acclimated to the southern culture, but never to the southern weather—he would be soaking wet with sweat before he got to the lab from his dorm room—but by the end of the summer, he had not only performed some original research but had organized the lab, and assigned specific projects to all the graduate students.

After the urging of many of the Ole Miss Physics faculty, Ron returned to Oxford and entered graduate school in physics in 1981. It was soon obvious that Ron was an exceptional student. After he finished his MS degree, and having won an Office of Naval Research (ONR) Graduate Student Fellowship, Ron returned to the Northeast and the ivied halls of Yale University where he joined Bob Apfel's research group. His colleagues at Yale, Christy Holland and Carr Everbach, among many others, recall his tireless efforts to manage Apfel's cluttered laboratory and make progress on his Ph.D. research project: "The characterization of individual cell properties by ultrasonic scattering." Ron was also a compassionate mentor to his fellow graduate students, and was given the honor of teaching acoustics to both undergraduate and graduate students.

After graduation from Yale, Ron stayed on to run Bob Apfel's start-up company, Apfel Enterprises, Inc., exploring his entrepreneurial bent. Somehow the South had grown on Ron, so he returned to Oxford as one of the first employees of the National Center for Physical Acoustics. Ron was involved in a number of research projects, but soon set about trying to explain the acoustics associated with bubbles in the ocean. Ron conducted a multi-institutional experiment at Lake Seneca in New York to determine the characteristics of bubble cloud scattering. The experiment was difficult and initially failed; however, Ron was persistent and succeeded in the unique measurement of low frequency scattering from a submerged bubble cloud, the only free field measurement of collective scattering, which was published in the *Journal of the Acoustical Society of America* (*JASA*) in 1992. This experimental effort exhibited Ron's attributes: the ability to work arduous hours, his considerable experimental abilities, and his dogged persistence to see a project through to its conclusion.

In 1991, Ron decided to try out the West Coast for a change and joined the Applied Physics Laboratory (APL) in Seattle, where together with Larry Crum, they formed an acoustics research group at the University of Washington (UW). At APL/UW, Ron was involved in research in sonoluminescence, ambient noise in the ocean, acoustic cavitation

in liquids, and a number of widely diverse topics. Throughout his entire career, Ron is known for the rigor of his experimental studies and his insight into all sorts of physical phenomena. His publications soon attracted the attention of Allan Pierce who was building a strong acoustics group at Boston University and Ron circled back to the East coast. At Boston University Ron's work on sound scattering from microbubble distributions including the dynamics of bubbles, mutual bubble interactions and the properties of bubbly liquids had significant impact on underwater acoustics and sonar signal processing.

In 2006, Ron was selected as the 65th George Eastman Professor of Oxford University, which is a distinguished Chair appointed annually to a citizen of the United States who is deemed to be of great eminence in teaching or research in any field of study. Prior Eastman Professors include Deans, Distinguished Faculty Chairs, a Justice of the U.S. Supreme Court, and 12 Nobel Prize winners. Ron was the first engineer to hold this post in the long history of this prestigious award. Following his year in Oxford, he returned to Boston University where he is currently the Chair of the Department of Mechanical Engineering.

Ron has been an active member of the Acoustical Society of America, rarely missing a meeting, serving on numerous committees, including the Executive Council, and perhaps having the singular distinction of being the only individual who has served as Chair of two Technical Committees. With nearly 40 *JASA* articles and almost 100 others in similar journals, Ron is considered one of the world's experts in biomedical ultrasound, acoustic cavitation, and bubble physics.

Ron is married to Nancy Stone Roy and they have two children, Caitlyn and Sydney, both Boston University graduates.

LAWRENCE A. CRUM
WILLIAM M. CAREY
ALLAN D. PIERCE

Acoustical Society of America
GOLD MEDAL

Jiri Tichy

2010

The Gold Medal is presented in the spring to a member of the Society, without age limitation, for contributions to acoustics. The first Gold Medal was presented in 1954 on the occasion of the Society's Twenty-Fifth Anniversary Celebration and biennially until 1981. It is now an annual award.

PREVIOUS RECIPIENTS

Wallace Waterfall	1954	Lothar W. Cremer	1989
Floyd A. Firestone	1955	Eugen J. Skudrzyk	1990
Harvey Fletcher	1957	Manfred R. Schroeder	1991
Edward C. Wente	1959	Ira J. Hirsh	1992
Georg von Békésy	1961	David T. Blackstock	1993
R. Bruce Lindsay	1963	David M. Green	1994
Hallowell Davis	1965	Kenneth N. Stevens	1995
Vern O. Knudsen	1967	Ira Dyer	1996
Frederick V. Hunt	1969	K. Uno Ingard	1997
Warren P. Mason	1971	Floyd Dunn	1998
Philip M. Morse	1973	Henning E. von Gierke	1999
Leo L. Beranek	1975	Murray Strasberg	2000
Raymond W. B. Stephens	1977	Herman Medwin	2001
Richard H. Bolt	1979	Robert E. Apfel	2002
Harry F. Olson	1981	Tony F. W. Embleton	2002
Isadore Rudnick	1982	Richard H. Lyon	2003
Martin Greenspan	1983	Chester M. McKinney	2004
Robert T. Beyer	1984	Allan D. Pierce	2005
Laurence Batchelder	1985	James E. West	2006
James L. Flanagan	1986	Katherine S. Harris	2007
Cyril M. Harris	1987	Patricia K. Kuhl	2008
Arthur H. Benade	1988	Thomas D. Rossing	2009
Richard K. Cook	1988		

CITATION FOR JIRI TICHY

. . . for contributions to acoustical intensity measurement, active noise control, education in acoustics, and for service to the Society

BALTIMORE, MARYLAND • 21 APRIL 2010

Jiri Tichy was born in the former Republic of Czechoslovakia. He studied electrical engineering and physics, received a doctorate in Technical Sciences (DSc) from the Czech Technical University in Prague in 1952, and was awarded the Diploma Habil in 1965. Later, he was appointed Associate Professor of Physics in the Department of Physics at the Czech Technical University. Early on he was interested in acoustics and published many papers on acoustics topics.

Jiri was acquainted with two American acousticians: Lawrence Batchelder and Leo Beranek who helped him to immigrate to the United States in 1968. He landed at the Pennsylvania State University, teaching architectural acoustics in the Department of Architectural Engineering and the Acoustics Graduate Program. His interaction with the large Penn State acoustics community, spread over six academic departments, and the Applied Research Laboratory allowed him to expand his professional work in acoustics. He was promoted to full professor in 1972 and was transferred to the Graduate Program in Acoustics where he was appointed the chairman of the Program. That program, which was initially an intercollege graduate program, was expanded later to become a full-fledged graduate department in the College of Engineering. It is the only program in the United States that awards M.S. and Ph.D. degrees in acoustics. Jiri served as the Program Head until his retirement in 1997, serving for 25 years.

Jiri's research expertise was primarily in audio engineering, noise, acoustic intensity, and active noise control. He is a pioneer in the measurement of acoustic intensity techniques, and has written seminal papers on the topic. These include "Measurement of the Complex Acoustic Intensity and the Acoustic Energy Density," with G. Elko in *Proceedings of Inter-Noise 84*, and "Instantaneous and Time-Averaged Energy Transfer in Acoustic Fields," with J. Mann and A. Romano in the *Journal of the Acoustical Society of America* (*JASA*) in 1987. He presented a Tutorial Lecture on this subject at the 115th meeting of the Acoustical Society of America (ASA) and delivered keynote speeches at the International Seminar on Noise Control, Rio de Janeiro, Brazil in 1989 and the Acoustical Society of Japan meeting in 1997.

Jiri Tichy's pioneering contributions in active noise control include the use of feed-forward control in signal processing techniques for the control of noise from many sources, within enclosures, rooms, halls, and ducts. Two of his seminal contributions are "Global Attenuation of Broadband Noise Fields Using Energy Density Control," with Y. C. Park and S. D. Sommerfeldt in *JASA* in 1997, and "Narrowband and Broadband Active Control in an Enclosure Using the Acoustic Energy Density," with J. W. Parkins and S. D. Sommerfeldt in *JASA* in 2000. He delivered keynote speeches on the subject at the tenth Brazilian Conference of Mechanical Engineers (1989), the Acoustical Society of Korea meeting in 1991, Acoustical Congress of Chile, as well as Internoise '95 and Internoise '01. These keynote speeches delivered at international meetings indicate the remarkable reputation and stature Jiri holds in acoustics in the national and international arenas.

Jiri has been a member of ASA since he came to the United States. He has devoted a major part of his entire professional society activities to ASA. He is a Fellow of the Society and has served in many positions. These include Chair of the Technical Committee on Noise, member of the Technical Council, Technical Program Chair for the ASA meeting in State College (1997), and Co-organizer and Co-chair of the ASA/EAA Meeting in Berlin (1999). He served very effectively as Vice-President of ASA (1992) and as its President (1993). He also served on various committees, such as publication policy, education, and medals and awards.

He has served as President of the Institute of Noise Control Engineering (INCE) (1986), and its Board of Governors (78-83), General Chair of Noise-Con '87, and '97, and General Chair of ACTIVE '95 and '97. He is also active in the Audio Engineering Society and is an Honorary Member of the Czech Acoustical Society. Such activities involving national and international societies point to a highly respected and acclaimed scientist and leader in acoustics.

Jiri Tichy has dedicated his life to education in acoustics. He has nurtured the Acoustics Graduate Program at Penn State since 1972. During those years, graduate enrollment increased from 30 to over 100 graduate students. He established a Continuing Education Program for industry, and a Certificate Program in Acoustics. He also mentored 28 M.S. and Ph.D. candidates in acoustics throughout his tenure at Penn State.

Jiri and his wife Dagmar, who passed away in 2003, were married in Prague in 1955. They had one daughter who is currently the head of the Oncology and Hematology Department at the Toledo Children's Hospital, Ohio. Jiri and his present wife, Iva, split their time between Toledo and Prague.

In conclusion, I am very happy that the Society has recognized Jiri's many accomplishments by awarding him the prestigious Gold Medal

SABIH I. HAYEK

Session 4aAAa

Architectural Acoustics and Engineering Acoustics: Rooms for Reproduced Sound I

K. Anthony Hoover, Chair
McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362

Invited Papers

9:00

4aAAa1. Design of critical listening rooms: A historical perspective. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

Critical listening room design has progressed to follow loudspeaker evolution from mono to two-channel to multi-channel, as well as the advances in loudspeaker quality. There are basically two fundamental approaches to two-channel room design. The non-environment room, in which the contribution of the room is minimized, and versions of live-end-dead-end, in which early reflections were minimized creating a spatio-temporal reflection free zone and diffuse reflections from the rear of the room are used to create passive surround sound. This presentation will focus on an immersive design for multi-channel loudspeaker formats, which utilizes broad bandwidth absorption and diffusion. Research with low-frequency plate resonators and a broad bandwidth diffusion chamber has contributed to the proposed design and results will be discussed. The proposed multi-channel critical listening room includes the use of low-frequency control down to 40 Hz provided by 4-in.-thick metal plate resonators on a portion of the front wall and in all corners, broad bandwidth diffusion on all wall and ceiling surfaces, and strategically placed multiple subwoofers. The intent is to minimize monophonic reflections and create lateral enveloping energy by uniformly scattering sound from all of the active sound sources in the room.

9:20

4aAAa2. Work on the reproduction of sound at the Visualization Laboratories at the King Abdullah University of Science and Technology: A case study. Steve Ellison (Meyer Sound Labs, Inc., Berkeley, CA 94702, ellison@meyersound.com) and Peter Otto (Sonic Arts, CalIT2/UCSD, La Jolla, CA 92093, potto@ucsd.edu)

Researchers at the newly opened King Abdullah University of Science and Technology are developing new ways to visualize and auralize complex data sets in immersive environments. These environments include the Interactive Media Room, a 100-seat multi-purpose facility, and Cornea, a $10 \times 10 \times 10$ in.^{3} six-sided stereoscopic cave. Both facilities are equipped for multi-channel audio generation and playback in various standard and custom formats, as well as electronically variable acoustics utilizing electroacoustic architecture. Concurrent viewing and dialogue between participants in the two spaces is supported. Fully immersive environments present unique challenges due to physical properties of projection screens and the geometries that characterize these structures. Explorations of Cornea at KAUST and StarCave at UCSD are presented, along with strategies for reproducing sound and varying acoustics in these environments. Multi-use facilities such as the IMR also have conflicting acoustical goals. For instance, the optimal acoustic for multi-channel or cinematic reproduction is less reverberant than optimized classroom environments. The use of electroacoustic architecture to improve the listening and overall immersive experience at both Cornea and the Interactive Media Room is discussed. Examples are presented including an experimental VR emulation of the acoustics of the King Abdullah Grand Mosque on the KAUST campus.

9:40

4aAAa3. Seelos Theater: A case study for a renovated multipurpose room with multiple loudspeaker systems. Matthew J. Moore, Alexander G. Bagnall, and Aaron M. Farbo (Cavanaugh Tocci Assoc., 327 F. Boston Post Rd., Sudbury, MA 01776, mmoore@cavtocci.com)

The Seelos Theater at Holy Cross College has been renovated twice since starting out as a bowling alley in the main student center on campus. This most recent renovation provided an opportunity for the school and design team to update and modernize this heavily used lecture hall and cinema. The authors will show how two different loudspeaker systems and AV systems were integrated with the architectural acoustics design built for reproduced sound. The authors found that the resultant performance of the systems performed better than was expected in computer modeling. Existing mechanical system noise mitigation, integration of existing film equipment, short construction timelines, and serving the needs of different audience types will be discussed. The authors used lessons learned with this project to help with the design of spaces for future projects.

10:00

4aAAa4. Speaker coverage in active acoustic music practice rooms. Ronald Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55060, ron.freiheit@wengercorp.com)

Larger music practice rooms (>25 m^3) present unique challenges compared to smaller, traditional practice rooms (6–12 m^3) when employing active acoustics technology and integrated digital recording. An experiment was conducted comparing a ceiling array of 32 speakers in a large music practice room ($5.3 \times 5.7 \times 2.7$ m^3, 81 m^3) with that of eight speakers located in corners of that room (one at the

top and one at the bottom in each corner facing parallel to the walls). The goal was to determine which speaker configuration would be the most effective for creating even sound field coverage and uniform frequency response. Data were collected for both seated and standing positions in the room from a matrix of 42 locations. Frequencies from 40 Hz to 8 kHz were plotted for analysis. The data were compared to the qualitative preferences expressed by a number of musicians using the rooms for both practice sessions (with the active acoustics enabled) and listening to the playback of recorded practice sessions.

10:20

4aAAa5. Small music venues and amplified sound: High-sound pressure levels and architecture. David S. Woolworth (Oxford Acoust., Inc., 356 CR 102, Oxford, MS 38655, dave@oxfordacoustics.com)

This paper is a survey of typical popular live music venue sound pressure levels observed in Oxford, MS. An attempt is made to measure clarity and articulation in the venues with increasing sound pressure, and the results are correlated with previous observations of Knudsen and others. Additional analysis of challenges to sound operators and programs to control high-stage volumes are discussed.

10:40

4aAAa6. Mixing studio isolation. Anthony K. Hoover (McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Westlake Village, CA 91362)

A new suite of mixing and teaching studios had been discussed for many years at Berkley College of Music in Boston. The authorization to proceed with design and construction of the studios finally came with a new Dean of the Music Technologies Department, who, without benefit of much of the background and previous discussions, contacted an old colleague for the final studio design, and then contracted with a reputable full-service contractor for construction services. Some unusual problems developed during construction, which sensitized the client to subsequent situations, in turn requiring acoustical consulting services. This paper will review some of the project history, outline the problems, and discuss solutions and recommendations, and will address some of the concerns with the specialized construction and unusual complications.

11:00

4aAAa7. Edison phonograph recording and reproduction in a concert hall. Alex U. Case (Sound Recording Tech., UMass Lowell, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

A recent recording session at the University of Massachusetts Lowell merged the birth of sound recording with the state of the art. A small ensemble performed an original piece of music—composed in the style of early 1900s American music—on stage in Durgin Concert Hall. It was recorded live to both high-definition digital multitrack and an original Edison wax cylinder phonograph. As the acoustical influence of the phonograph is substantial, several takes were needed to adjust the balance of the live performance to maximize the musical impact of the reproduced sound realized through wax cylinder playback. Multiple sonic viewpoints were simultaneously recorded: stereo, surround sound, close microphones, and Edison phonograph. The resulting recordings are played and discussed.

11:20

4aAAa8. Soul spaces: How acoustics influence the music production process and the recordings that result. Matthew B. Zimmern (8 Illsley Hill Rd., West Newbury, MA 01985, matthewzimmern@gmail.com)

Spaces in modern recording studios are quite different than those of 40 years ago. While today it is common for recording facilities to have various multi-purpose, acoustically designed spaces, many recordings from the middle of the 20th century were often captured in a single room, one that was not constructed with acoustical principles in mind and often times featured acoustical deficiencies. Soul music production practices of the late 1960s were researched, and an original composition was recorded in two different ways: by using past recording techniques and by using contemporary musical practices. Through listening to audio examples of the same song recorded in two different spaces, the sonic qualities of the different recording environments will be discussed, as well as how the different production techniques affected the recording, musical, and aesthetic attributes of the resulting audio recordings.

11:40

4aAAa9. Obtaining a frequency independent reverberation time in listening rooms. Niels Werner Adelman-Larsen (Flex Acoust., SCION-DTU, Diplomvej 377, 2800 Kgs. Lyngby, Denmark, nwl@flexac.com), Cheol-Ho Jeong, and Jiazi Liu (Tech. Univ. of Denmark, 2800 Kgs. Lyngby, Denmark)

High-frequency sound emitted from loudspeakers is very directive and can be easily aimed at the listeners. Investigations show that even a few people decrease the mid-/high-frequency reverberation time of a room significantly both due to actual absorption but also due to the scattering of the sound that they introduce which makes the placement of absorptive material important. People absorb approximately one-sixth of the sound at low frequencies than at high frequencies, but still much reproduced music contains very loud levels of bass frequency sound. In order to avoid loud, reverberant bass sounds masking the direct mid-/high-frequent sound, the room needs to be designed with quite large amounts of bass absorption in order to properly serve its purpose. Helmholtz and porous type absorbers need a large cavity in order to also achieve bass absorptive qualities and this is a challenge in many smaller rooms. Membrane absorbers require less space and due to the omni-directional behavior of bass sound their placement is not so critical. This paper will present new measurements of absorption coefficients of standing persons as well as a new type of modular bass absorber system: 2.6 in. deep, absorption coefficient of 0.6 from 50–125 Hz, and designed to readily incorporate into the architectural design.

Session 4aAAb

Architectural Acoustics: Hidden Gems

Andrew N. Miller, Chair
BAi, LLC, 4006 Speedway, Austin, TX 78751

Chair's Introduction—8:00

Invited Papers

8:05

4aAAb1. The Festival Hill concert hall. Richard Boner and Pamela Harght (BAi, LLC, 4006 Speedway, Austin, TX 78751 rboner@baiaustin.com)

The Festival Institute at Round Top, TX, was established in 1971, and began with ten pianists performing on an open stage, under the stars. Following the dream of the founder, Concert Pianist James Dick, Festival Hill has grown to be a premier summer venue for classical music, with over 30 concerts performed each June and July. The August-to-April Series, the International Guitar Festival, the Theatre Forum, The Poetry Forum, and the Herbal Forum bring the total number of year-round events to more than 50. This paper focuses on the excellent and visually unique 1000 seat concert hall, which was built in stages over a 26-year period from 1981–2007. Also discussed are the renovated 19th-century homes which have been moved to Festival Hill, serving as housing and practice spaces for musicians.

8:30

4aAAb2. Whispering arches as intimate soundscapes. E. Carr Everbach (Swarthmore College, Swarthmore, PA 19081) and David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683)

Whispering archways are whispering galleries. They are sometimes found around smooth arches, doorways, and even bridges. Persons positioned on opposite sides of a whispering arch communicate privately by whispering into the arch. In that way they function as private and even intimate soundscapes. Some whispering arches that have stood for centuries have inspired local folk legends. Such acoustical architecture can help to establish a sense of place and strengthen emotional ties. It is said that a whispering arch at the entrance to a 10th c. Gothic cathedral at the historic monastic site of Clonmacnoise in the Irish midlands figured in local courtship for centuries. There, shy young lovers are said to have uttered vows of love and proposals of marriage. The arch was also used by lepers to give confession without infecting the priest. Whispering arches were almost certainly unintended in the past but can also be created intentionally as an intimate form of public art. Recent measurements on a whispering archway at West Chester University show that high background noise levels from traffic or air handlers can ruin an otherwise intimate space; however, so modern architects should plan the entire soundscape carefully.

Contributed Papers

8:55

4aAAb3. A performers guide to venue acoustics. David J Zartman (Graduate Program in Acoust., Penn State Univ., 405 EES Bldg., State College, PA 16801)

Whether in high school or the Vienna Symphony, musicians have little control over their performance venues. An instrument can be upgraded or replaced if there is an undesirable aspect, but replacing a venue is often far too expensive to be a tenable solution. A musician's perfect venue will also change piece to piece—an incredible venue for lush Wagner operas can cause issues for someone performing an intricate Paganini concerto and vice versa. What then are these effects? Why do they take place? How can a performer anticipate and combat undesirable aspects of venues while drawing out their featured characteristics for an optimized performance?

9:10

4aAAb4. Vineyard shape variations of the Concertgebouw, Amsterdam. Weihwa Chiang (Natl. Taiwan Univ. of Sci. and Technol., 43, Keelung Rd., Section 4, Taipei 106, Taiwan, Republic of China) and Wei Lin (Hwa Hsia Inst. of Technol., Chung Ho, Taipei, Taiwan, Republic of China)

The Concertgebouw, Amsterdam is a unique one of its kind because of the extra width and size for a classical rectangular hall and the significant portion of seats behind the stage. These characteristics make the hall as "semi-surround" as the vineyard halls of the present day. Computer simulation was performed to analyze the possibility of deriving design variations out of the Concertgebouw. Clarity, strength, sectional balance, and stage support of various settings of stage position and overall proportion were compared as the first step. Variations that contain the geometrical features of vineyard shape halls were further developed to optimize the three acoustical qualities and to reduce the differences among seats. The architectural features evaluated included surface treatments, arrangement of individual seating blocks, and the geometry of upper walls and the ceiling.

Session 4aAAc

Architectural Acoustics and Physical Acoustics: Physical Acoustics in Boston Symphony Hall: A Guide for the Perplexed

James B. Lee, Chair
6016 SE Mitchell, Portland, OR 97206

Chair's Introduction—9:25

Invited Papers

9:30

4aAAc1. Back to vitruvius? James B. Lee (6016 S. E. Mitchell, Portland, OR 97206, cadwal@macforcego.com)

There is no generally accepted theory of acoustics of concert halls based on the physics of sound as waves. Lord Rayleigh, founder of the science, remarked but little about sound in rooms. Wallace Clement Sabine made an analogy of sound in rooms to the kinetic theory of gases; although this is theoretically untenable, Sabine designed the superb concert hall which is the subject of this session. Leo Leroy Beranek published the first systematic examination of concert halls, demonstrating that subjective evaluations of acoustic quality are consistent: everyone agrees which are good and which are poor; nonetheless Beranek encountered severe difficulty in applying his experimental results. What aspects of the physics of Rayleigh explicate the data of Beranek and the design of Sabine?

9:55

4aAAc2. Simple acoustic source radiation near a large wall. Michael J. Moloney (Dept. of Phys. and Optical Eng., Rose-Hulman Inst. of Technol., 5500 Wabash Ave., Terre Haute, IN 47803, moloney@rose-hulman.edu)

An acoustic simple source is predicted to radiate twice as much sound near an infinite rigid wall as it does in free space. As the simple source becomes more distant from the wall, its radiated power is predicted to decrease in an oscillatory way toward the value it has in free space. This behavior can be observed indirectly by fitting a 10-s signal of decaying amplitude from a tuning fork resonator box at various distances from a large unobstructed wall. Each fit determines the damping constant b, which reflects power losses, including radiated acoustic power. In a plot of b versus distance from the wall, the oscillatory part of the damping constant closely matches the theoretical curve of total radiated power from a simple acoustic source near an infinite rigid wall.

10:20—10:30 Break

10:30

4aAAc3. Evaluation of a boss model and subtraction technique for predicting wideband scattering phenomena in room acoustics. Georgios Natsiopoulos (WSP Acoust., Rullagergatan 4, 41526 Gothenburg, Sweden, georgios.natsiopoulos@wspgroup.se)

Scattering from a hard hemisphere on an infinite plane is studied experimentally and theoretically in a laboratory environment using a subtraction technique. With this technique, the scattered field is determined by subtracting the transient responses from the plane with and without the boss present. Scattering for different combinations of transducer and hemisphere positions are then classified into a number of cases, denoted in-plane/off-plane and back-/forward/mixed scattering. Theoretical and experimental results are compared and evaluated from an auralization point of view using a one-third octave band smoothing of the contribution from the hemisphere. With the smoothing used, average intensity level discrepancies typically less than 3 dB are obtained for a six octave wide ka-interval, where a equals the hemisphere radius.

10:55—11:00 Break

11:00—12:00 Panel Discussion

Session 4aAB

Animal Bioacoustics, Signal Processing in Acoustics, and Noise: Topical Meeting on Signal Processing for Subtle and Complex Acoustic Signals in Animal Communication I: Automated Classification of Animal Acoustic Signals I

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Sean K. Lehman, Cochair
Lawrence Livermore National Lab., Livermore, CA 94550

Chair's Introduction—8:55

Invited Papers

9:00

4aAB1. Model-based signal processing approach to animal bioacoustics: A brief overview. James V. Candy (Lawrence Livermore Natl. Lab., P.O. Box 808, L-151, Livermore, CA 94551)

The model-based approach to signal processing is generally founded on the fundamental concept of incorporating any *a-priori* knowledge of the underlying phenomenology from which the signal evolved along with measurement instrumentation and uncertainty (noise, parameters, etc.) in the form of mathematical models that are embedded in the processor. In this way, the phenomenologist, experimenter, and signal processor combine all of their possible knowledge into a scheme enabling each to think within their own comfort zones while developing a powerful approach to extract the illusive information they desire. In this overview, we present the concepts required to develop mode-based processing schemes that can be used in a wide variety of animal bioacoustics applications ranging from signal estimation, tracking, identification, detection, and classification. We discuss the development of this approach incorporating acoustic applications that can be extrapolated to animal bioacoustics problems. We can express all of these techniques in terms of a model-based framework enabling the use of such powerful estimation techniques such as Kalman filters (Gaussian case) to Bayesian particle filters (non-Gaussian case) for solution to a wide variety of problems.

9:20

4aAB2. Detecting and identifying cetaceans species from their acoustic emissions. Whitlow Au and Julie Oswald (Hawaii Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734)

In the past decade, the use of autonomous remote passive acoustic recorders to detect the presence of cetaceans has grown enormously throughout the world. Since these devices digitize acoustic signals at sampling rates in the tens of kHz, a single device can quickly accumulate many Gbytes of data in a short time. Therefore, automatic detection and recognition algorithms must be developed to handle the large amount of data. There is no general signal method processing method that can be applied to all species of cetacean. The algorithm by the name of ROCCA (real-time odontocete call classification algorithm) developed by Dr. Julie Oswald is probably the most general one that has been applied to the whistles of dolphins. The program, extensible bio acoustics tool (XBAT) developed by Harold Figueroa, is probably the most general tool for use with baleen whales. However, investigators also tend to develop their own algorithms to suit their requirements. The problems and difficulties in developing automatic detection and recognition algorithms will be the focus of this presentation. General principles associated with the geography, biology, and the characteristics of sound emissions in relation to the development of algorithms will be discussed along with examples of different approaches.

9:40

4aAB3. Overcoming the idiosyncrasies of spectrogram correlation detection. Kurt M. Fristrup (Natural Sounds Program, Natl. Park Service, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

Spectrogram correlation has been widely used to detect animal sounds. Effective application of this method to identify Black-capped Vireo (*Vireo atricapillus*) sounds in 22 000 h of environmental recordings required the use of multiple spectrogram templates for each general type of sound to ensure that substantial subsets of these sounds were not missed by the detector. Over 5 million candidate sounds were identified, but the majority of these were not confirmed by expert review. Accordingly, acoustical features were extracted from a stratified random sample of 8284 candidate sounds that were identified by experts, and a random forest classifier was developed to winnow out the false alarms. The estimated classification error varied by site from 0.5% to 6%. When this classifier was applied to the entire data set, approximately 740 000 vireo sounds were identified. This project illustrates both the feasibility of monitoring birds over large spatial and temporal scales and the challenge of adequately sampling the range of variation in a very restricted class of biological signals.

10:00—10:15 Break

10:15

4aAB4. Individual identification using hidden Markov models for population monitoring and assessment. Michael T. Johnson (EECE, Marquette Univ., 1515 W. Wisconsin, Milwaukee, WI 53233) and Patrick J. Clemins (Assoc. for the Advancement of Sci., Washington, DC 20005)

Individually distinct acoustic features are present in a wide range of animal species, just as they are in humans, and the widespread success of speaker identification in humans suggests that robust automatic identification of individual animals from their vocalizations is attainable. Despite this, only a few studies to date have yet attempted to use individual distinctiveness to help assess population structure, abundance, and density patterns. Here we present an approach, based on individual identification and clustering using hidden Markov models (HMMs), which enables a more direct mechanism for using individual vocal variability to monitor and assess populations. Current results indicate that the new method is able to give good estimates of local abundance based on vocalization clustering, which can in turn be used in an acoustic mark-recapture framework to estimate population. Limitations to this approach currently include the need for explicit call-type separation prior to individual clustering, which is possible in many species but can create a problem in species with unknown or variable repertoires. Overall, it is hoped that this new technique may lead to a more accurate understanding of population structure and abundance on a larger scale.

10:35

4aAB5. Identifying delphinid whistle contours using graph search. Marie A. Roch, Bhavesh Patel (Dept. of Comp. Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-7720), Shannon Rankin (NOAA SW Fisheries Sci. Ctr., La Jolla, CA 92037-1022), Yvonne Barkley (Bio-Waves Inc., Encinitas, CA 92024), Melissa S. Soldevilla (Duke Univ., Beaufort, NC 28516), and John A. Hildebrand (Univ. of California, San Diego, La Jolla, CA 92093-0205)

The automated characterization of delphinid whistle contours can lead to insights (both potential and realized) into biological questions such as habitat use and behavior. Prior to the 1990s, most measurements of whistle contours were conducted manually by trained analysts, and even today the noisy signal environment offers significant challenges to fully automated classification systems. In this talk, we provide an overview of challenges and historical approaches to this problem. We present our recent research using graph search algorithms to characterize complex auditory scenes involving numerous animals vocalizing simultaneously. We end by discussing techniques that have the potential to further advance this area of research. [Work sponsored by ONR.]

10:55

4aAB6. Wavelets: A comparison with the spectrogram and other methods for time-frequency analysis. Patrick Loughlin (Dept. of Bioengineering, Univ. of Pittsburgh, 745 Benedum Hall, Pittsburgh, PA 15261, loughlin@pitt.edu) and Leon Cohen (Hunter College, CUNY, New York, NY)

Many natural signals exhibit spectral content that changes over time. Methods for time-varying spectral analysis first emerged in the 1940s with the development of the "sound spectrograph" at AT&T Bell Laboratories. Since then, the spectrogram has become the primary method for time-frequency analysis. Originally implemented as a bank of band-pass filters, today the spectrogram is typically computed digitally via the short-time Fourier transform. Recently, wavelets have been proposed as a superior method for time-frequency analysis. The usual argument is that the spectrogram uses a fixed window length, whereas the wavelet approach uses windows that are longer for lower frequencies and shorter for higher frequencies. While the benefits of this approach are usually taken as self-evident, we explore in critical detail the aims of time-frequency analysis, and the benefits afforded by wavelets versus the spectrogram and modern methods such as the Choi–Williams distribution. In particular, since a primary aim of time-frequency analysis is to study the local spectral and temporal characteristics of signals, we examine the local moments of the various methods. Local moments are related to important signal features such as the instantaneous frequency and bandwidth. We show the effect of fixed windowing versus variable wavelet windowing on these features.

11:15—12:00 Demonstrations and Discussion

Session 4aAO

Acoustical Oceanography and Underwater Acoustics: Acoustic Inversions in Ocean Environments

Yong-Min Jiang, Cochair
Univ. of Victoria, School of Earth and Ocean Science, P.O. Box 305, Victoria, BC V8W 3P6, Canada

Alexandra I. Tolstoy, Cochair
ATolstoy Scientific, Inc., 1538 Hampton Hill Cir., McLean, VA 22101

Contributed Papers

8:30

4aAO1. Variation of uncertainty and resolution with problem formulation in continuous geoacoustic inversion. Andrew A. Ganse and Robert I. Odom (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th, Seattle, WA 98105, aganse@apl.washington.edu)

In continuous geoacoustic inversion, the resolution and uncertainty of the estimated ocean bottom are inherently linked by a tradeoff with each other. They also depend on the experiment geometry and the choice of formulation for the data and for the bottom model. For example, the resolution and variance will differ if the data are represented by an intensity envelope timeseries or a tau-p timeseries, or if wave slowness or bottom impedance is estimated. Previous work by the authors investigated variations in resolution and uncertainty with geometry and problem formulation based on linearization of the geoacoustic problem at a given solution. However, the problem is further complicated by the fact that the solution point itself can also depend on the geometry and data and model formulation so that the resolution and uncertainty vary for that reason also. This aspect of the nonlinear geoacoustic inverse problem is explored here. [Work supported by ONR.]

8:45

4aAO2. The estimation of geoacoustic parameters via frequencies 25–100 Hz. A. Tolstoy (ATolstoy Sci., Inc., 1538 Hampton Hill Cir., McLean, VA 22101)

This work will discuss recent efforts to extend a previously discussed low-frequency (LF) geoacoustic inversion method to slightly higher frequencies. In particular, the earlier method was "successful" at frequencies 25–50 Hz where an exhaustive search was possible even in the presence of errors in other (assumed known) parameters. However, some of the test data analyzed offered only one appropriate frequency (53 Hz) and did not converge to a unique solution. In fact, thousands of data fits were found. Here, we shall examine efforts to improve convergence by allowing for slightly higher frequencies. Range may also be a consideration at these LFs.

9:00

4aAO3. Travel time inversion of broadband data from Shallow Water 2006 experiments. Yong-Min Jiang (minj@uvic.ca) and N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, Canada)

This paper presents geoacoustic inversions of broadband signals collected by the L-shaped array SWAMI32 during the Shallow Water 2006 Experiments. The L-shaped array was deployed in 70 m of water off the coast of New Jersey. The vertical leg of the array (VLA) has 10 even spaced sensors which expends 53.55 m in the water column, while the horizontal leg (HLA) has 20 uneven spaced bottom moored sensors that give 256.43 m of the aperture. An acoustic source was maintained at a depth of 35 m and towed along a circle around the VLA at a speed of 0.5 knots. The distance between the acoustic source and the VLA was around 190 m. Mid-frequency (1100–2900-Hz) chirps received at the VLA and HLA were analyzed for investigating the variability of the sea bottom properties around the circle. The data at the HLA were analyzed to assist the identification of the bottom layer structure while the data at VLA were employed to carry out the geoacoustic inversion. Environmental data collected in the vicinity were used in the inversion to account for the variable water column environment. [Work supported by ONR Ocean Acoustics.]

9:15

4aAO4. Low-frequency attenuation and sound-speed measurements in marine sediments using an impulsive source. Altan Turgut, Jeff Schindall, and Steven Means (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

Low-frequency (200–1000-Hz) sediment attenuation and sound-speed measurements were conducted at a site on the New Jersey Shelf where additional data sets from a chirp-sonar bottom profiler (2–12 kHz) and an acoustic-probe system (10–80 kHz) are available. Impulsive sound signals, generated by automated light-bulb implosions, are received by a 16-element vertical line array at short ranges ($<$500 m). Precursor arrivals and signals reflected from the R-reflector are used to estimate sediment sound-speed and attenuation. Attenuation in dB/m/kHz is estimated using the spectral-ratio technique and sound-speed is estimated from the travel-time analysis. Frequency dependency of sound-speed and attenuation is also investigated within a wide frequency band (200 Hz–80 kHz) using the results from impulsive source, chirp-sonar, and acoustic-probe measurements. Measured attenuation and sound-speed values seem to be well predicted by an extended Biot theory for sediments with distributed pore sizes. [Work supported by ONR.]

9:30

4aAO5. Studies on the effect of shear on compressional wave attenuation. Gopu R. Potty and James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

Results of a modeling study to examine the effect of shear on compressional wave attenuation are presented. This study also investigates a shear inversion algorithm based on interface waves using synthetic data. Recent studies suggest that inclusion of shear speed is necessary to explain the correct frequency dependence of attenuation. Synthetic data will be generated for elastic bottom, with different shear speeds, and these data will be inverted for compressional wave attenuation. This could provide insight into the effect of shear speed on the attenuation coefficient obtained from inversion at various frequencies. In addition to investigating this effect, we also develop inversion algorithms for shear speed. One of the most promising approaches is to invert the relation between seismo-acoustic interface waves (Scholte waves) that travel along boundaries between media and shear wave speed. The propagation speed and attenuation of the Scholte wave are closely related to shear-wave speed and attenuation over a depth of 1–2 wavelengths into the seabed. The dispersion characteristics of the Scholte wave has been successfully used for inversion of sediment shear properties. Synthetic data will be used in our study to develop an inversion scheme. [Work supported by Office of Naval Research.]

9:45

4aAO6. Toward passive acoustic remote sensing of ocean-bottom gas seeps. Chad A. Greene and Preston S. Wilson (Dept. Mech. Eng. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

Cold seeps are gas vents that are found in the ocean, often along continental shelves near sediment-borne methane hydrates. Methane hydrates and methane gas seeps are of particular interest both for their potential use as an energy source and for their possible contribution to global climate change. This work is an initial step toward passively locating cold seeps and quantifying their gas flow rates using acoustic remote sensing techniques. Results are presented from laboratory experiments in which gas fluxes were determined from the radiated acoustic signature of a model seep. The physical principle that supports the technique and its accuracy will be discussed. [Work supported by ONR.]

10:00—10:15 Break

10:15

4aAO7. The card-house structure of mud: Energy between particles and its effect on bubble formation. Joseph O. Fayton (Rensselaer Poly. Inst., Troy, NY 12180, faytoj@rpi.edu), Allan D. Pierce, William M. Carey (Boston Univ., Boston, MA 02215), and William L. Siegmann (Rensselaer Poly. Inst., Troy, NY 12180)

A model of mud as a lyophobic colloid [Verwey and Overbeek (1948)] leads to a card-house structure of platelets, typically kaolinite or smectite particles. Because of isomorphous substitution, each platelet carries a distributed negative charge. The structure is immersed in water which, especially so for sea water, carries positive and negative ions, and the positive ions tend to cluster on both sides of the platelets, so that the composite platelet charge is zero, but each platelet is analogous to a sheet of longitudinal electrical quadrupoles. The charge distribution causes parallel platelets to repel each other, but which can attach so that one platelet edge can touch the face of another platelet. The energy associated with such attachments is discussed, and it is conjectured that it is associated with van der Waals and London attraction forces. Recent sound transmission experiments by Carey at Dodge Pond in Connecticut support the contention that sediment mud invariably contains vapor bubbles, and other work such as that of Boudreau *et al.* (2005) suggests that the bubbles are considerably flattened, with shapes that have been described as corn-flakes. Present paper conjectures as to whether these shapes are associated with the card-house structure of mud.

10:30

4aAO8. An investigation of the combustive sound source. Andrew R. McNeese, Jason D. Sagers, Preston S. Wilson (Dept. of Mech. Eng. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292), and David P. Knobles (The Univ. of Texas at Austin, Austin, TX 78713-8029)

The combustive sound source (CSS) is a versatile impulsive underwater sound source with an adjustable bandwidth and output amplitude. Unlike typical impulsive acoustic sources, CSS can maintain a wide bandwidth at low amplitude; hence, it is a more environmentally friendly impulsive source for ocean acoustics experiments and surveys. The CSS consists of a submersible combustion chamber, open to the water, which is filled with a combustible fuel/oxidizer mixture. The mixture is ignited and a combustion wave propagates through the mixture. During this process the ensuing bubble expands due to an increase in temperature and can collapse to a smaller volume than before ignition. This bubble activity radiates acoustic pulses. The CSS can be used as a source for low-frequency sediment characterization and TL measurements, and when deployed on the bottom can produce seismic interface waves. In addition to stationary deployments in the water column, CSS can be deployed in a tow body and as an array. Discussion will focus on the latest CSS design including functionality, acoustic output, long-term operational stability, and future development plans.

10:45

4aAO9. Detection and classification of sub-surface features based on their seismo-acoustic signatures. Irena Lucifredi and Miaki Ishii (Harvard Univ., 20 Oxford St., Cambridge, MA 02138, irenav@alum.mit.edu)

Complex subsurface structure makes seismo-acoustic imaging difficult, raising the need for novel detection, classification, and imaging approaches. For the representation of range-dependent seismo-acoustic propagation, we employ a hybrid, coupled wavenumber integration approach to range-dependent seismo-acoustic modeling based on the OASES environmental modeling framework. In feature identification and classification, time series analysis frequently provides the sought-after recognition clues. Nevertheless, the geometry and the structure of the subsurface features generate a variety of complexities to the wave field caused by different physical mechanisms, geometric constraints, and intrinsic properties. Based on the time series attributes such as the time of arrival, waveform amplitude, and the frequency content, a set of identification features is extracted and a set of corresponding classes is formed. Using the simulated data scenarios, general rules incorporating the identification features of subsurface structures are deduced. This classification methodology is then applied to the data recorded by an array of seismographs and hydrophones on and off the coast of Japan. The results create a basis for classification and description of the structures of interest such as the presence or absence of very low-velocity range-dependent layers below the water-seabed interface.

11:00

4aAO10. Numerical experiments on weak shock propagation in marine sediments. B. Edward McDonald (Naval Res. Lab., Washington, DC 20375, mcdonald@sonar.nrl.navy.mil)

NRL is planning a set of laboratory and field experiments to investigate weak shock propagation in saturated marine sediments. Questions to be addressed are as follows: (1) How much does the granular content of the sediment increase the effective nonlinearity coefficient? (2) Does the 3/2 order Hertzian nonlinearity inherent to grain contacts play a significant role? Numerical and analytic solutions involving the Hertzian nonlinearity reveal maximum nonlinearity coefficient near zero stress, whereas in a fluid, nonlinearity increases with stress. Numerical experiments are presented using the NPE model [McDonald and Kuperman, J. Acoust. Soc. Am. 81, 1406 (1987)] to determine whether shocks resulting from Hertzian nonlinearity can be observed in the presence of nominal values of frequency-linear attenuation common to granular media. [Work supported by the Office of Naval Research.]

11:15

4aAO11. Tank experiments for validation of volume scattering models. Jorge E. Quijano and Lisa M. Zurk (Dept. Elec. and Comput. Eng., Portland State Univ., 1900 SW 4th Ave., Ste. 160, Portland, OR 97201)

Mid-frequency scattering from random media is of interest for active sonar applications and, in particular, in shallow water environments, where ocean bottom interaction may affect the performance. To estimate the contribution of volume scattering from the sea bed, models based on the wave equation have been developed and more recently, an approach based on transport theory was proposed [J. Acoust. Soc. Am. **126**, 1711–1723 (2009)]. Validation of this model is difficult to achieve in field experiments due to environmental uncertainties and in many cases to the lack of ground truth revealing the structure of the random media. As an alternative, tank experiments were implemented in a controlled environment at scaled frequencies to explore the range of applicability as a function of different input parameters such as frequency of operation, scatterer size, and concentration. For this work, gel slabs containing random distributions of micron-sized glass beads were manufactured, and broadband pulses at the frequency band 200–600 kHz were utilized to characterize the scatterer contribution. Acoustic propagation through this heterogeneous media is analyzed using the proposed radiative transfer method and the results are compared to the analytical solution of the wave equation for the long-wavelength approximation.

11:30

4aAO12. Inverting for surface displacement fields using directly measured point-to-point sensitivity kernels. Jit Sarkar, Shane Walker, Bruce Cornuelle, William A. Kuperman (Scripps Inst. of Oceanogr., UCSD, Mail Code 0238, 9500 Gilman Dr., La Jolla, CA 92093-0238, jit@mpl.ucsd.edu), Philippe Roux, and Christian Marandet (LGIT, BP 53, 38041 Grenoble Cedex, France)

The effect of surface perturbations on underwater acoustic fields has been directly measured in an ultrasonic tank experiment. High-frequency transducer arrays of 64 elements each are placed 600 mm apart, submerged in ~50 mm of water. A point perturbation is used to displace the surface sequentially in 1-mm steps between the two arrays. At each position a 3.8-MHz signal is transmitted from the source array (round-robin), and the response recorded on the receive-aray, both with and without the presence of the surface probe. The difference between these two measurements yields the point-to-point acoustic sensitivity to that perturbation. These data are explored with respect to inferring the surface structure through both linear full-wave inversions and double-beamforming.

THURSDAY MORNING, 22 APRIL 2010 GALENA, 7:55 A.M. TO 12:05 P.M.

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Induced Cellular Bioeffects

E. Carr Everbach, Chair
Swarthmore College, Dept. of Engineering, 500 College Ave., Swarthmore, PA 19081-1397

Chair's Introduction—7:55

Invited Papers

8:00

4aBB1. Temporary modulation of vascular barriers with focused ultrasound and microbubbles. Nathan McDannold (Dept. of Radiology, Brigham & Women's Hospital, 221 Longwood Ave., Rm. 521, Boston, MA 02115)

The combination of focused ultrasound and preformed microbubbles (ultrasound contrast agents) circulating in the bloodstream provides an opportunity to utilize mechanical stimulation of endothelial cells to temporarily and locally modulate transvascular transport and permeability. This stimulation may directly alter the endothelium and increase free transport of agents out of the vasculature or it may trigger a physiological response and increase active transport. This talk will review recent work in the brain and kidney utilizing low-intensity focused ultrasound bursts and microbubbles to temporarily disrupt barriers to transvascular passage. In the brain, these sonications result in a temporary disruption of the blood-brain barrier and potentially enable a means for targeted drug delivery. Recent work will be shown demonstrating that the method is sensitive to anesthesia agent, perhaps due to differences in vasoactive effects. In the kidney, we have found that these sonications result in a temporary increase in glomerular filtration rate and urine production and the ability to temporarily pass larger molecules, potentially providing a new platform for the study, and perhaps treatments, of renal disease. Overall, the interaction between ultrasound, microbubbles, and the endothelium presents a unique means to interact with blood vessels and provides opportunities to develop novel treatments.

8:20

4aBB2. Study of sonoporation at the single cell level and cellular bioeffects associated with sonoporation facilitated by microbubbles. Cheri X. Deng (Dept. of Biomedical Eng., Univ. of Michigan, 2200 Bonisteel Blvd., Ann Arbor, MI 48109)

The interaction of ultrasound-driven microbubbles with cells results in multifaceted cellular bioeffects including physical disruption of cell plasma membrane and subsequent downstream effects. Microbubble facilitated sonoporation, the ultrasound-induced disruption of plasma membrane, enhances the transport of ions, other intracellular contents, and external agents through the membrane into the cytoplasm via diffusion, thereby making it useful for intracellular drug and gene delivery. Transport of these entities, which depends on the dynamic process of pore formation and resealing, plays important roles in many downstream cellular effects of sonoporation beyond the initial pore formation and subsequent diffusion-related transport. For example, extracellular calcium ions diffused into the cytoplasm initiate the process of membrane resealing. In addition, intracellular calcium transients are generated which may be related to many cellular processes triggered by the important second messenger molecule. Recent results of sonoporation at the single cell level and related bioeffects will be discussed. Electrophysiological techniques reveal the dynamic process of sonoporation. Real-time fluorescence imaging measurements of intracellular calcium concentration in mammalian cells subjected to sonoporation demonstrate the spatiotemporal evolution of sonoporation related calcium transients including the large influx of calcium in sonoporated cells and the complex dynamic calcium oscillations and waves. [Work supported by NIH.]

8:40

4aBB3. Photoacoustic delivery of molecules into cells. Mark R. Prausnitz, Prerona Chakravarty, Wei Qian, and Mostafa A. El-Sayed (Georgia Inst. of Technol., Atlanta, GA 30332, prausnitz@gatech.edu)

A major barrier to drug and gene delivery is crossing the cell's plasma membrane. This study presents a physically based mechanism to deliver molecules into cells with high efficiency and viability using photoacoustic effects generated by carbon nanoparticles. The results demonstrated intracellular delivery using this method in multiple cell types for uptake of small molecules, proteins, and DNA. At optimized conditions, calcein uptake was seen in up to 50% of cells with nearly 100% viability and in 90% of cells with ⩾90% viability. Uptake was not observed when cells were irradiated in the absence of carbon nanoparticles. These studies suggest that uptake occurs due to transient membrane permeabilization resulting from explosive photoacoustic forces generated by laser-induced carbon-steam reaction, C(s) + H2O(*l*) = CO$(*g*) + H2(*g*). This synergistic use of nanotechnology with advanced laser technology could provide an alternative to viral and chemical-based drug and gene delivery.

9:00

4aBB4. Ultrasound standing wave fields induce endothelial cell sprouting within three-dimensional engineered tissue. Kelley A. Garvin (Dept. of Biomedical Eng., 205 Goergen Hall, Univ. of Rochester, Rochester, NY 14627, garvin@bme.rochester.edu), Denise C. Hocking, and Diane Dalecki (Univ. of Rochester, Rochester, NY 14627)

The field of tissue engineering is working to develop fully functional replacement tissues and organs. To achieve this goal, methodologies aimed at controlling the growth of new vascular systems in three-dimensional (3-D) engineered tissues are needed. We hypothesized that organizing endothelial cells into multicellular, planar bands of cells within 3-D collagen gels using the radiation forces developed in an ultrasound standing wave field (USWF) would promote an angiogenic endothelial cell phenotype. Human umbilical vein endothelial cells were suspended in an unpolymerized type-I collagen solution and were exposed to continuous wave USWFs. The collagen solution was allowed to polymerize during the 15 min USWF treatment to maintain the USWF-induced banded pattern of cells within a 3-D collagen gel. Following a 24 h incubation period, endothelial cell sprouts were observed emerging from USWF-induced endothelial cell bands. The average length of these sprouts was ~100 μm. Sprouting was absent in sham samples where a rounded cell morphology was observed. The influence of acoustic exposure parameters on endothelial cell sprouting was investigated. These studies indicate that USWF technologies promote formation of capillary precursors in 3-D engineered tissue and thus, this technology has the potential to advance the field of tissue engineering.

9:20

4aBB5. Histological evaluation of continuous and pulsed high-intensity focused ultrasound induced vascular bioeffects in the presence or absence of contrast agents. Anna Tokarczyk, Ian Rivens, and Gail ter Haar (Joint Dept. of Phys., ICR: Royal Marsden NHS Foundation Trust, 15 Cotswold Rd., Sutton SM2 5NG, United Kingdom, anna.tokarczyk@icr.ac.uk)

A novel experimental model which uses an isolated, cannulated rat mesenteric artery (~ 400 μm diameter) has been used for the visualization of vessel behavior and damage during therapeutic ultrasound (US) exposures in the presence of contrast agents. The experimental setup includes a fluorescence microscope, ultrasound transducer, and a chamber in which the vessel, attached to micropipettes, can be perfused with a feeding buffer ± contrast agents and fluorescent dyes. A range of continuous and pulsed 1.7 MHz high intensity focused ultrasound (HIFU) exposures have been used at therapeutic and sub-therapeutic intensities (pressures). Vessel wall damage and leakage of intravascular buffer have been observed, predominantly in the presence of contrast agents. The observed lack of vascular constriction and expansion in response to phenylephrine and acetylcholine indicates smooth muscle and endothelial cell damage. This has been confirmed by histology (haematoxylin and eosin staining) and immunohistochemistry. Factor VIII and alpha-actin antibodies (which have receptors localised on endothelial, and smooth muscle cells respectively) have been used for accurate localisation of HIFU affected areas. This model will be also used to investigate the bio-effects induced by the exposure of vessels to diagnostic US exposures in the presence of contrast agents.

9:40

4aBB6. A pre-treatment planning strategy for high-intensity focused ultrasound treatments. Billy Andre (Dept. Mech., Boston Univ., 110 Cummington St., Boston, MA 02215, billy.andre@gmail.com), Phillip Jason White, Nathan McDannold, and Gregory L. Clement (Brigham and Women's Hospital, Boston, MA 02115)

High-intensity focused ultrasound (HIFU) therapy is a non-invasive treatment method for uterine fibroids patients. However, there exists focal distortion, due to inhomogeneous tissues, that can divert therapeutic dosage to non-targeted locations. A missing component of current HIFU protocols is a simulated pre-treatment planning, a procedure that can predict the efficiency of ultrasonic energy delivery pathways based on MRI scans using a simulated propagation model. To conduct this study, a 1.502-MHz transducer transmitted HIFU through five *ex vivo* bovine tissue specimens in an MR-compatible tank setup. A needle hydrophone (aperture = 0.2 mm) scanned the acoustic pressure field orthogonal to the axis of propagation for three different positions of the tissue specimens (0, 4, and 8 deg, about the rotation mounts). To quantify the level of distortion, the distortion index (DI = 1—ratio between local and total acoustic energy fields) were calculated. The data collected confirmed HIFU distortion due to tissue, where 48% of the focal acoustic intensity was redistributed away from the focal field. Focal distortion decreased on average by 7.2% (ranging from 5.6% to 12.3%, SD = 2.8%) upon optimizing the transducer-specimen orientation the default transducer orientation (0 deg). In addition, the propagation simulation model accurately predicted HIFU distortion from a specimen's MRI data.

10:00—10:30 Break

10:30

4aBB7. The harmonic motion imaging for focused ultrasound system for tumor detection and treatment: Simulation, *in vitro* and *in vivo* results. Elisa Konofagou, Caroline Maleke, and Yi Hou (Dept. of Biomedical Eng., Columbia Univ., 1210 Amsterdam Ave., New York, NY 10023, ek2191@columbia.edu)

Harmonic motion imaging (HMI) for focused ultrasound (HMIFU) is a novel, all-ultrasound-based system that can simultaneously detect and localize tumors as well as subsequently generate and monitor their ablation based on their distinct stiffness. In this paper, we present the results of a fundamental simulation study that is then validated using *in vitro* and *in vivo* findings. The same HMI and HMIFU parameters as in the experimental studies were used, i.e., the high-intensity focused ultrasound (HIFU) frequency was 4.68 MHz and the modulation frequency equal to 25 Hz. An HIFU-simulator was used to predict the lesion and its resulting time-dependent displacement field. Using the same parameters, *in vitro* bovine liver experiments were performed to validate the size of the simulated lesions. A transgenic mouse model of invasive adenocarcinoma was used for *in vivo* feasibility using the same parameters. Good agreement between the simulated lesion maps and pathology findings of the *in vitro* liver experiments was found. The lesion formation was identified by a 30% decrease in displacement amplitude in vivo. Tumor cell death was also confirmed by histology. Based on these results, the HMIFU system may offer a cost-efficient and reliable alternative for real-time monitoring of thermal ablation.

Contributed Papers

10:50

4aBB8. Controllable plasma membrane poration by pulsing tandem microbubble. Georgy Sankin, Fang Yuan, and Pei Zhong (Dept. of Mech. Eng. and Mat. Sci., Duke Univ., Box 90300, Durham, NC 27708)

Understanding the dynamic interaction of cavitation bubbles with biological tissue is central to the effective and safe application of therapeutic ultrasound in clinical medicine. Coupled oscillation of two laser generated microbubbles (maximum radius = 28 μm) in constrained media and associated shear stresses are investigated experimentally. Bubble-bubble interaction in a microchannel of 25-μm height is observed using high-speed video cameras and μPIV technique. Two liquid micro-jets moving in opposite directions can be generated when the second bubble is produced at the maximum size of the first one. The interaction of these tandem microbubbles with single cell leads to controllable poration of adjacent cell membrane and dye uptake. Micro-PIV data are compared with cell viability at various bubble-cell distances and azimuthal orientations. This method provides a new approach for highly selective cell treatment *in situ* applicable to targeted microinjection of macromolecules and gene vectors in microfluidics devices. [This work was supported in part by NIH.]

11:05

4aBB9. Ultrasound induced mechanical induction of mesenchymal stem cells. Jia-Ling Ruan, Yak-Nam Wang, and Stuart B. Mitchell (1013 NE 40th St., Seattle, WA 98105)

Low-intensity pulsed ultrasound (LIPUS) has been used to accelerate fracture healing and tissue regeneration but the biological mechanism of these responses is not completely understood. Stem cell activity can be induced through biochemical and mechanical mechanisms. Despite the common use of LIPUS in fracture healing and tissue regeneration, there are only a few studies that examine the mechanical induction of stem cells with ultrasound. The purpose of this study is to determine the effects of ultrasound-generated mechanical stimulus on the behavior of human mesenchymal stem cells (hMSCs) *in vitro*. In our preliminary studies low-intensity pulsed ultrasound was used to induce mechanical strains on hMSCs *in vitro*. Amplitudes, pulse durations, and pulse repetition frequencies were varied such that different radiation pressures were generated on hMSCs in culture. Results indicated a significant increase in cell proliferation after 4 consecutive days of treatment as well as a significant difference in the cellular response between treatment parameters. Results suggest that LIPUS can be used to influence mechanical mediated stem cell behavior; however, more research is needed to fully elucidate the relationship between ultrasound and hMSC response.

11:20

4aBB10. Local tissue-sparing due to blood flow in medium-size vessels. Simon Woodford, Ian Rivens, Gregory Vilensky, Nader Saffari, and Gail ter Haar (Dept. of Phys., Inst. of Cancer Res., Sutton, Surrey SM2 5NG, United Kingdom)

High-intensity focused ultrasound (HIFU) is a noninvasive surgical technique that uses ultrasound energy deposition to thermally destroy selected tissue. However, the cooling effect of blood flow can lead to unwanted tissue-sparing during HIFU treatment, particularly in the tissue immediately adjacent to the vessel wall. Regions where large (clinically detectable) vessels are present will require an increase in intensity or treatment time. However, there is a range of vessel sizes that are thermally relevant yet clinically undetectable. The treatment protocol must be chosen such that the spared region around these vessels is minimized or eliminated completely. Using numerical simulations, we have determined preliminary recommendations for suitable treatment protocols.

11:35

4aBB11. Dynamics and flow field produced by coupled oscillation of tandem microbubble. Fang Yuan, Georgy Sankin, Jaclyn Lautz, and Pei Zhong (Dept. of Mech. Eng. and Mat. Sci., Duke Univ., Durham, NC 27708)

Coupled oscillations of two adjacent laser-induced microbubbles have been shown to produce unique asymmetric bubble deformation and microjet formation. The resultant microstreaming and shear stress can cause localized cell membrane poration with potential application in targeted drug and gene delivery. In this study, we investigate the bubble dynamics and flow field produced by laser-generated tandem microbubble in a microfluidic device. Flow field around the tandem microbubble is analyzed with respect to phase delay, inter-bubble distance, and size ratio between the two microbubbles. In addition, micropatterning technique is used to control the adhesion site and growth pattern of HeLa cells in relation to the tandem microbubble. Flow vorticity is observed to be a key parameter that correlates with the strength of tandem microbubble oscillation and resultant macromolecule uptake efficiency. [Work supported by NIH Grant Nos. R01DK052985, R21CA135221, and S10RR016802].

11:50

4aBB12. Ultrasound-mediated nail drug delivery system to treat fungal disorders. Danielle Abadi and Vesna Zderic (Dept. of Elec. and Comp. Eng., The George Washington Univ., 801 22nd St. NW, Washington, DC 20052, zderic@gwu.edu)

This physiotherapy device treats nail fungal disorders by improving drug delivery to the nail bed using low frequency ultrasound to increase the permeability of the nail. Applying therapeutic ultrasound to the nail will allow pits to form on the surface through which more of the topical drug can be delivered. Transducers with corresponding matching networks are constructed in-house from 300-kHz and 1-MHz piezoelectric crystals. A software interface allows patients to select which toes to treat and at what intensity level (high, medium, or low). The highest-intensity setting supplies 1.5 W/cm^2 for a duration of 30 s to prevent damage to tissue below the nail. The software interface sends information to the electrical driving system,

which is composed of an amplifier for each intensity level and a signal generator. A Franz diffusion cell setup in combination with a spectrophotometer is used for experimental testing of nail permeability by measuring the amount of drug delivered through the membrane. Animal nails are used as a human nail model, and both drug-mimicking dye and Penlac (a topical prescription drug) are used for testing.

THURSDAY MORNING, 22 APRIL 2010 LAUREL A/B, 8:00 TO 11:45 A.M.

Session 4aEA

Engineering Acoustics: Sound Projection and Transduction

Stephen C. Thompson, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Chair's Introduction—8:00

Contributed Papers

8:05

4aEA1. On the concept of transient directivity: Pipes and horns. Daniel Tengelsen, Brian E. Anderson, and Timothy W. Leishman (Dept. of Phys. and Astron., Acoust. Res. Group, Brigham Young Univ., Provo, UT 84602)

Directivity is a convenient way to represent how sound radiates from some arbitrary object in steady-state. The steady-state condition is implied when time harmonicity is assumed. Because all physical systems do not begin in steady-state, this directivity measurement is only valid after the transient portion of the solution has decayed. Transient directivity—a measure of sound radiation versus angle at a given instant in time and before the system has reached steady-state—is presented. Understanding an object's transient and steady-state radiation characteristics may be important in understanding the sound radiation from sources that are transient in nature. Results of transient directivity will be presented for pipes and horns from both numerical models and experimental measurements.

8:20

4aEA2. Modeling of transducer arrays for direct digital-to-analog conversion of signals. Jose Amado, Nikita Tkachov, and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, jose_amado@student.uml.edu)

A binary weighted array of speakers will be used to reconstruct a decomposed sequence of delta functions. The idea of directly converting digital signals to an analog acoustic output was first proposed by J. L. Flanagan in 1979, in which he designed, fabricated, and tested digital transducers for 4-, 5-, and 6-bit PCM signals. Flanagan found that at 6-bit resolution, the system fell short of good quality, and that condenser transducers had a limited output sound level of about 85 dB. Simulations will be used to investigate experimentally developed models for transducers and apply the direct digital-to-analog approach.

8:35

4aEA3. A new loudspeaker design for the enhancemenet of sound image localization on flat display panels. Gabriel Pablo Nava, Keiji Hirata, and Yoshinari Shirai (NTT Commun. Sci. Labs., NTT Co., Hikaridai 2–4, Seikacho, Kyoto 619-0237, Japan, pablo@cslab.kecl.ntt.co.jp)

In most audio-visual multimedia applications, conventional stereo loudspeakers have been used to implement auditory displays. However, a fundamental problem with this kind of displays is that only the listeners situated at the sweet spot and over the symmetrical axis of the loudspeaker array are able to accurately localize the sound images. Although a number of audio signal processing algorithms have been proposed to expand the listening area, relatively less study on new loudspeaker configurations has been explored. This paper introduces a simple, yet effective, loudspeaker design to enhance the localization of sound images over the surface of flat display panels. In contrast to previous approaches, expansion of the listening space is achieved by attachment of rigid barriers which physically modify the sound radiation pattern of the loudspeakers. Moreover, numerical simulations, experimental sound measurements, and subjective tests have been preformed to validate a prototype of the proposed loudspeaker design using a display panel of an immersive teleconferencing system. Finally, an example of an interactive application was implemented involving real-time speaker tracking with a microphone and video cameras.

8:50

4aEA4. Head-tracking interface using a Wii remote. Megha Sunny, Ayse Kalkan-Savoy, and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, megha_sunny@student.uml.edu)

In this work, we will examine the problem of detecting the angular motion of head in effort to build a head-tracking system to control sound. The head motion occurs in *X*, *Y*, and *Z* linear directions and three rotation angles, namely pitch, roll, and yaw. In the past, we had difficulties to detect the rotation angles, especially yaw. Our current work focuses on the detection of rotation angles using a software interface program. The distance between sound source and head, and angular rotation data is collected by the Wiiremote's built-in optical sensor and three-axis accelerometer. These real-time data will be used as input to our software interface. Results will be used in conjunction with head related transfer function to create the three dimensional sound source effects.

9:05

4aEA5. The influence of matching layer material loss on radiation impedance conversion in ultrasonic transducers. Minoru Toda (Measurement Specialties Inc., 135 Gedney Rd., Lawrenceville, NJ 08648, minoru.toda@measspec.com)

PZT based thickness mode ultrasonic transducers for both air and water /tissue typically have a quarter wavelength front matching layer with impedance *Zm*. It is widely known that the lower acoustic impedance Z_R of the propagation medium is converted to a higher impedance at the PZT surface by the relation $Z\max = Zm^2/Z_R$ (quarter wavelength conditions). In this work, the converted impedance was accurately calculated using a transmission line model incorporating the mechanical quality factor *Qm* of the matching layer material. In a medical transducer, it was found that the peak value is 20% or 28% lower than $Z\max = Zm^2/Z_R$ for $Qm = 15$ or 10, respectively. For air transducers, the peak value is one to two orders lower than *Z*max for the same range of *Qm*. These *Qm* values are typically ob-

served for the filled epoxy matching layer materials used in medical transducers and also for porous or air entrained materials used for air acoustic transducers. A simple impedance conversion equation for an air transducer has been proposed, making the design of air impedance matching layers easier and suggesting that neglecting material loss leads to serious errors.

9:20

4aEA6. Transitory response of an acoustic levitator. Sahir Shakir and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, sahir_shakir@student.uml.edu)

This work will examine the transitory response of an acoustic levitator. Simulations will be used to ascertain the relationship between viscous and inertial forces on the vertical conveyance of a solid projectile. The objective is to use standing waves of strong intensity in a cavity to elevate a solid projectile. The projectile is suspended by nonlinear acoustic forces, and by rate of change in the frequency, amplitude, or cavity length, the transitory response will be determined.

9:35

4aEA7. Two driving constructions of loudspeakers for low-frequency range by piezoelectric ultrasonic motors. Juro Ohga (Ohga's Acoust. Lab., 2-24-3, Tamanawa, Kamakura, 247-0071, Japan), Takehiko Adachi, Hiroki Saito, Ryosuke Suzuki, Gen Takeda, Hajime Kubota (Chiba Inst. of Technol., Narashino 275-0016, Japan), Hirokazu Negishi (MIX Acous. Lab., Shiba, Minato-ku, Tokyo 105-0014, Japan), Kazuaki Maeda (TOA Corp., Takarazuka 665-0043, Japan), and Ikuo Oohira (Oohira's Lab., Aobadai, Aoba-ku, Yokohama 227-0062, Japan)

The loudspeaker driven by piezoelectric ultrasonic motors is characterized by a precise very-low-frequency reproduction due to its high-driving mechanical impedance. It has a lot of merits comparing to the conventional electrodynamic loudspeakers. One of the reason will be that this loudspeaker is a power flow modulator, not a transducer. In this presentation, two sorts of ultrasonic motors are compared as driver elements of the loudspeakers. One is an ordinary revolution-type motor and the other is a reciprocal linear motion type actuator. The authors constructed and improved practical low-frequency-range loudspeakers by using continuous revolution of ultrasonic motors. Its final model uses combination of two motors with same axis, which drive two cone radiators moving oppositely. This model shows a satisfactorily large output sound pressure and stable reproduction. However, its complicated elements for connection of the motors and the cone radiators cause a mechanical weakness. The authors, therefore, propose another new, completely different construction to avoid this defect. It applies linear motion of two ultrasonic actuators. A moving piece driven by piezoelectric ultrasonic vibrators is connected directly to a cone radiator. Comparison at various viewpoints and practical performance of these two constructions are presented at the meeting.

9:50

4aEA8. An ultrasonic vibrator constructed from laminated Galfenol steel. Scott P. Porter, Stephen C. Thompson, and Richard J. Meyer, Jr. (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA, 16804)

An ultrasonic vibrator has been developed to serve as the drive mechanism for an electroacoustic transducer. This design explores the unique characteristics of Galfenol, a recently invented giant magnetostrictive material. In addition to possessing competitive strain capabilities, strong mechanical properties, and a high-magnetic permeability, Galfenol does not require a prestress mechanism and can be laminated to effectively mitigate eddy current losses. Designing the vibrator required the authors to carefully engineer the magnetic circuit so that proper bias fields could be established using a permanent magnet. This step will be demonstrated with one- and two-dimensional models. Drive coil considerations will also be discussed and the fabrication and assembly of the vibrator will be shown along with in-air measurements.

10:05—10:15 Break

10:15

4aEA9. A preliminary analog circuit model of a balanced-armature transducer utilized for energy harvesting. Holly A. Smith and Stephen C. Thompson (Grad. Prog. in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, has202@psu.edu)

This research investigates a balanced-armature transducer's potential to harvest ambient vibrational energy into electrical energy that could then be used to power small devices or recharge batteries. Such a device is desirable due to its compact size and environmentally friendly operation. An analog circuit model of a balanced-armature transducer was created in a version of SPICE. To ensure the model's accuracy, the electrical impedance predicted by SPICE was compared to experimental measurements. The model was then adjusted for energy harvesting. [Work supported in part by the Office of Naval Research.]

10:30

4aEA10. Modified single crystals for high-power, broadband underwater projectors. Nevin P. Sherlock and Richard J. Meyer, Jr. (Appl. Res. Lab., P.O. Box 30, State College, PA 16804)

When operating high-power electromechanical devices, the performance is often limited by self-heating within the active material. This is especially true in high-power, broadband underwater projectors using the piezoelectric single crystal $Pb(Mg_{1/3}Nb_{2/3}O_3$–$PbTiO_3$ (PMN-PT). Although PMN-PT crystals show an excellent piezoelectric response ($d_{33} > 1500$ pC/N) and high-coupling coefficient ($k_{33} > 0.90$), device performance is limited by the high-mechanical losses ($Q_M < 100$) and low-temperature phase transformation ($T_{RT} = 95$ °C). This work describes the material property enhancements of two compositional modifications and compares the performance of these crystals in high-power, broadband transducers. One such modification is the addition of $Pb(In_{1/2}Nb_{1/2}O_3$ (PIN) or $PbZrO_3$ (PZ) to the binary PMN-PT composition. The greater thermal stability of ternary PIN-PMN-PT and PZ-PMN-PT crystals is shown by comparing the dielectric permittivity, piezoelectric coefficient, and coupling coefficient to unmodified PMN-PT as a function of temperature. Additionally, the mechanical losses of PMN-PT have been decreased by doping with $Mn^{3,4+}$ ions. Using a laser Doppler velocimeter, the losses are evaluated under increasing ac electric drive. Using these data, high-power, broadband projectors were constructed from these modified crystals, and the results are compared to a projector using unmodified PMN-PT. [Funded by ONR under N00014-07-1-0336.]

10:45

4aEA11. Modeling thermal mitigation and nonlinear behavior in high-power single crystal 1-3 composite transducers. Tara L. Tubbs and Richard J. Meyer, Jr. (Appl. Res. Lab., State College, PA 16804)

The desire for high-precision sonar systems has forced 1-3 composite transducers to the forefront of sonar design. Single crystal, which has improved mechanical and dielectric properties over PZT, provides a variety of advantages, and their implementation into a 1-3 composite design makes them a great candidate for sonar transducers. Driving these transducers at high levels, we can get a broad bandwidth and high-power in a smaller device. However, high power introduces overheating and nonlinear behavior in the material properties. Using finite element software GID/ATILA from ISEN along with COMSOL MULTIPHYSICS, it is possible to incorporate these phenomena and solve thermal mitigation problems. This allows for improved high-power single crystal 1-3 composite transducers.

11:00

4aEA12. Use of compressively stressed zinc oxide to increase microspeaker response. Lukas Baumgartel (Dept. of Phys., USC MEMS Res. Group, 3737 Watt Way, PHE 621, Los Angeles, CA 90089, lbaumgar@usc.edu) and Eun Sok Kim (USC MEMS Res. Group, Los Angeles, CA 90089)

A micromachined piezoelectric speaker was fabricated on a 5×5-mm^2, 1-μm-thick silicon nitride diaphragm. A 4×4-mm^2 zinc oxide (ZnO) piezoelectric transducer sits in the middle of the diaphragm, providing

actuation. Two variations were fabricated: one with the compressively stressed ZnO covering the region between the transducer and diaphragm perimeter—causing wrinkling—and another with the ZnO removed in this region. In both variations, the stress gradient causes curvature in the active area, raising the resonant frequency to above 4 kHz. The displacement response is therefore approximately flat from 40 Hz to 4 kHz. The speakers are driven with a sinusoidal voltage, and the response is measured with a laser interferometer. The wrinkled device exhibits 11 times larger response and can be actuated by much smaller voltage, achieving lower THD while still having a larger deflection. The wrinkled device is driven at 2 $V_{0.t.p.}$ from 40 Hz to 4 kHz, demonstrating a response of 55 nm/$V_{0.t.p.}$ and an average THD of 5.1%. The unwrinkled device is driven at 15 $V_{0.t.p.}$ over the same range, yielding a response of 5 nm/$V_{0.t.p.}$ and an average THD of 8.6%. Measured sound output and displacement spectra match each other well.

11:15

4aEA13. Analysis and design of a MEMS (microelectromechanical system) directional microphone diaphragm with active Q control. Ronald N. Miles, Quang T. Su, Weili Cui, Dorel Homentcovschi (Dept. of Mech. Eng., Binghamton Univ., Binghamton, NY 13902-6000, miles@binghamton.edu), and N. Eva Wu (Binghamton Univ., Binghamton, NY 13902-6000)

The analysis and design of a MEMS directional microphone are described that incorporates electronic feedback to achieve active Q control. The microphone diaphragm consists of a 1 × 2-mm stiffened plate fabricated out of polycrystalline silicon that is supported on a central hinge. The sound pressure gradient incident on the diaphragm produces a rocking motion about the central hinge. Interdigitated comb fingers at each end of the diaphragm enable both capacitive sensing and electrostatic actuation. The diaphragm has been designed to have its dominant resonant mode have a frequency of approximately 1 kHz. By minimizing sources of passive damping, the thermal noise of the microphone has been shown to be lower than the noise floor of existing two-microphone systems used in directional hearing aids [Miles *et al.*, J. Acoust. Soc. Am. **125** (2009)]. However, this low-passive damping also results in an undesirable resonance within the audible frequency range. To minimize the adverse effects of this resonance, a simple analog electronic feedback system is designed that can result in acceptable performance in both the frequency and time domains. [Work funded by NIH Grant R01 DC009429.]

11:30

4aEA14. Response of a MEMS (microelectromechanical systems) directional microphone diaphragm with active Q control. Quang T. Su, Ronald N. Miles, Weili Cui (Dept. of Mech. Eng., Binghamton Univ., Binghamton, NY 13902-6000, quang.su@binghamton.edu), Mihir Shetye (Solteras, City of Industry, CA 91748), and N. Eva Wu (Binghamton Univ., Binghamton, NY 13902-6000)

Measured results are presented that demonstrate the use of proportional and derivative electronic feedback to improve the performance of a directional microphone. The microphone diaphragm consists of a 1 × 2 mm stiffened plate fabricated out of polycrystalline silicon that is supported on a central hinge. The sound pressure gradient incident on the diaphragm produces a rocking motion about the central hinge. Interdigitated comb fingers at each end of the diaphragm enable both capacitive sensing and electrostatic actuation. The sound pressure gradient near the diaphragm has been measured by numerically differentiating the pressure measured by a probe microphone at locations around the diaphragm. The sound-induced motion of the diaphragm was measured using a laser vibrometer. From these measurements, estimates of the mechanical parameters of the diaphragm were obtained. By applying a known quasi-static voltage across the interdigitated fingers and measuring the resulting diaphragm deflection, an estimate for the derivative of the capacitance with respect to the displacement is obtained for the comb fingers of the diaphragm. Using these experimentally determined parameters and a linearized dynamic model of the system, the measured response of the microphone system with feedback is accurately predicted. [Work funded by NIH Grant R01 DC009429.]

THURSDAY MORNING, 22 APRIL 2010 DOVER C, 8:40 A.M. TO 12:00 NOON

Session 4aED

Education in Acoustics and ASA Committee on Diversity: Diversity Issues in Education in Acoustics

Juan I. Arvelo, Cochair
Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Preston S. Wilson, Cochair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin, TX 78712-0292

Chair's Introduction—8:40

Invited Papers

8:45

4aED1. "Future faces of physics" and other initiatives to broaden participation in science. Catherine O'Riordan and Kendra Rand (American Inst. of Physics, One Physics Ellipse, College Park, MD 20740, coriorda@aip.org)

Together, Hispanic-Americans and African-Americans make up over 25% of the US population, but they earn only 7% of physics bachelor's degrees. In order to help broaden the participation of underrepresented groups in STEM fields, the American Institute of Physics has several programs to work with students as well as to reach the general public. To engage physics undergraduates on the challenging subject of diversity, the Society of Physics Students (SPS), a society of over 4000 undergraduate physics students that is part

of the American Institute of Physics (AIP), has launched two new programs. SPS has developed a Physics Jeopardy-like game as part of a "Future Faces of Physics" kit that includes various materials that help engage students in conversations about diversity. These kits have been used at more than 20 regional and national meetings since 2008 with positive results. AIP has also worked with Research Corporation to launch a program that brings Nobel laureates to HBCU's and to SPS campus or regional events that include participation by Minority Serving Institutions. Another AIP program creates short TV science news segments and distributes them free of charge in Spanish. These programs and future directions will be summarized.

9:05

4aED2. What can we learn from statistics on acoustics? Rachel Ivie (Statistical Res. Ctr., American Inst. of Physics, One Physics Ellipse, College Park, MD 20740, rivie@aip.org)

This talk will present current statistics on acoustics degrees and employment in acoustics. The data come from the National Science Foundation's studies of degrees awarded and from the American Institute of Physics' surveys of members of our ten Member Societies (including ASA). Because acoustics encompasses many different disciplines, these statistics will be compared to data from larger fields such as engineering, physics, and life and earth sciences. Although the numbers are very small in acoustics, data will be shown on the participation of under-represented minorities and women where possible. Finally, implications for increasing diversity in acoustics will be discussed.

9:25

4aED3. What faculty say, and convey, matters: Interactions with underrepresented students in science, technology, engineering, and mathematics. Sharon Fries-Britt (Univ. of Maryland, 2203 Benjamin Bldg., College Park, MD 20742, sfries@umd.edu)

Join us for this invited presentation as we examine students perceptions of their interactions with faulty. The participants in this study were primarily undergraduate students; however, approximately one-quarter of the participants were graduate students and post-doctoral research assistants. The entire sample consisted of students from a variety of postsecondary institutions including public and private, predominantly White, historically Black and Hispanic serving institutions from across the United States. This research is part of a larger study conducted over a 5-year period (2004–2009) with the National Society of Black Physicists (NSBP) and the National Society of Hispanic Physicists (NSHP). The majority are physics majors; however, some students are pursuing dual degrees in other STEM disciplines such as math, astronomy, and engineering. The findings of this study indicate that their interactions with faculty in the classroom and in advising session are critical. When those interactions are positive students benefit tremendously; however, in many instances they are negative and the interactions can cause barriers to their engagement in learning process and in how supported students feel pursuing science. Join us as we discuss some of the challenges and opportunities that these students encounter in their interactions with faculty.

9:45

4aED4. Diversity in physics: Whose problem is this? What can I do? Theodore W. Hodapp (Dept. Education and Diversity, American Physical Society, One Physics Ellipse, College Park, MD 20740, hodapp@aps.org)

Physics has one of the lowest participation rates for underrepresented minorities and women of all science, technology, engineering, and mathematics (STEM) fields. Things are improving for women and, while still not representative of the population, the trends are encouraging. Underrepresented minorities, however, have not been as fortunate. I will describe the current status of participation in physics and discuss new initiatives planned to address the lack of involvement at all levels in the field. In particular, I will describe a new program by the American Physical Society (APS) that aims to significantly increase the number of minorities who receive Ph.D.'s in physics. Actions you can take within your community, university, or workplace will be discussed. I anticipate a lively discussion during the panel that follows to bring forward good ideas and possible actions we can collectively take to improve participation of *all* people in physics and the technical workforce. The APS has partnered with leaders in the community to address these issues, and we hope that members of the Acoustical Society will find ways to work collaboratively to attend to these problems. This is our responsibility.

10:05—10:30 Break

10:30

4aED5. Diversifying physics: Legal context, education action. Daryl E. Chubin (AAAS Capacity Ctr., 1200 New York Ave., NW, Washington, DC 20005)

Science education and career development are vital for (1) driving innovation and economic strength, (2) US leadership in producing research & development and the personnel responsible for its renewal, and (3) shaping federal investments in what universities do and K-12 schools teach. Yet the nation faces a demographic challenge: by 2042, the population will be "majority minority." Minorities represent less than 7% of the nation's STEM workforce and are grossly underrepresented in undergraduate and graduate degrees awarded in STEM fields. In addition, the US is flagging in STEM degree production compared to the nations of Europe and Asia. Regardless of one's political beliefs, the nation must prepare more of its citizens for careers in science and engineering. This lecture focuses on the legal climate for increasing participation of underrepresented groups (women and minorities) in physics education and careers. Review of student- and faculty-centered programs and practices that have been "effective" in research universities face another challenged: Can they be made "legally sustainable"? How to make progress "on the ground" in the face of what the law calls—for disciplines such as physics and specialties such as acoustics—a "pipeline problem" will be discussed.

10:50

4aED6. Vassar College-Bronx Institute acoustics workshop for low-income, ethnic minority, urban high school students. David T. Bradley (Dept. Phys. and Astronomy, Vassar College, 124 Raymond Ave., Poughkeepsie, NY 12604-0745, dabradley@vassar.edu) and Angela M. Kelly (Lehman College, City Univ. of New York, Bronx, NY 10468)

Recent studies continue to show that low percentages of women and members of underrepresented ethnic groups are pursuing careers in science, technology, engineering, and math fields. Acoustics is no exception to this trend. As national demographics change more rapidly, and globalization transforms the way we approach the scientific process, the need for diversity becomes even more essential. The inclusion of women and minorities not only changes the questions being asked but also changes how those questions are answered. For acoustics to grow as a discipline, barriers to participation and success of members from underrepresented groups must be eradicated, and we must address the leakage in the science pipeline for students and professionals at all stages in their careers. One key area of concern is the recruitment and retention of K-12 students. This presentation will detail the joint efforts of the Vassar College Physics and Astronomy Department and the Bronx Institute at Lehman College to establish a hands-on, inquiry based acoustic workshop series for urban, low-income, underrepresented ethnic minority students from a collection of high schools in the Bronx borough in New York City.

11:10

4aED7. Reinventing diversity. Howard J. Ross (Cook Ross, Inc., 8630 Fenton St., Ste. 824, Silver Spring, MD 20910, howardr@cookross.com)

Despite all of the efforts to find strategies to improve the way organizations are addressing diversity, inclusion, and cultural competency issues, some are still finding their goals unrealized. In response to this, we conducted extensive research within the field and focused on how to reinvent diversity for the 21st century. Our research points to the need for three major paradigm shifts: (1) A movement from the classic United States-based approach, which focuses too heavily on race and gender and an assimilation model of diversity, to one that incorporates a deep understanding of globalism and the impact of major changes in population demographics around the world, global business, and interactive communication and networking; (2) a shift from the "good person/bad person paradigm" of diversity, which has developed and permeated a corrective mindset about diversity; (3) a 'find them and fix them' approach, which escalates the "us vs. them" way that people approach the issue and makes it more rather than less difficult to address. We have to move away from the event-based way for we have approached diversity, a pattern that has given us many specific activities, but not enough emphasis on systems thinking and culture-based change, to one that is strategic, systemic, and culture based.

11:30—12:00 Panel Discussion

THURSDAY MORNING, 22 APRIL 2010 GRAND BALLROOM III/IV, 8:30 TO 9:45 A.M.

Session 4aPAa

Physical Acoustics: Nonlinear Systems

Matthew E. Poese, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Contributed Papers

8:30

4aPAa1. Dynamic stabilization of the ultradamped inverted pendulum. Randy Carbo (Acoust., Penn State Univ., University Park, PA 16802, rmc258@psu.edu), Matthew E. Poese, and Robert W.M. Smith (Appl. Res. Lab., University Park, PA 16802)

The dynamic behavior of the damped inverted pendulum is analogous to the modes of fluid systems with nonzero viscosity where a denser fluid is situated above a lighter fluid (e.g., the Rayleigh–Taylor instability). Thermoacoustic devices have such configurations whenever a hot heat exchanger is situated below a cold heat exchanger. It has been shown that such systems can be stabilized by effectively modulating gravity and can be described by a damped Hill equation. A numerical solver was developed that uses a stroboscopic technique to determine whether solutions are stable (bounded) or unstable (unbounded). While it is sometimes assumed that the stability boundaries of undamped systems bound the damped cases, for large values of damping ($Q < 0.05$, where Q is the quality factor), the numerical solution predicts that damping can destabilize the system for certain regions of the parameter space. Results of an experiment performed by the authors using a physical pendulum and eddy current damping to verify this rather counter-intuitive result are described. [Research supported by the Office of Naval Research and ARL Exploratory and Foundational Research Program.]

8:45

4aPAa2. Molecular dynamics of nonlinear and nonequilibrium effects in sound propagation. Takeru Yano (Dept. of Mech. Eng., Osaka Univ., Suita, 565-0871, Japan, yano@mech.eng.osaka-u.ac.jp)

Nonlinear and nonequilibrium effects on the propagation process of large amplitude and high-frequency sound waves are studied with the method of molecular dynamics, where the whole physical phenomenon is described in the numerical solution of Newton's equation of motion for hundreds of thousands of gas molecules. The frequency of sound studied here is so high that the wavelength should be comparable with the mean free path of gas molecules, and hence the continuum theory cannot be applied to the resulting phenomenon. The wave profiles are obtained by averaging the molecular motions, and the most conspicuous nonlinear effect in the result is the mass and energy transports by a shock-like wave, for which the origin can be attributed to a nonequilibrium effect caused by the high-frequency sound. The nonequilibrium effects are quantitatively examined with the velocity distribution function of gas molecules.

9:00

4aPAa3. Application of the chaotic cavity transducer concept for imaging in non-reverberant media. Koen Van Den Abeele, Bart Van Damme, Steven Delrue (Wave Propagation and Signal Processing, K.U. Leuven Campus Kortrijk, E. Sabbelaan 53, B-8500 Kortrijk, Belgium, koen.vandenabeele@kuleuven-kortrijk.be), and Olivier Bou Matar (Ecole Centrale de Lille, Villeneuve dAscq 59652, France)

The concept of chaotic cavity transducer utilizes a combination of a piezoelectric ceramic disk glued on a cavity of chaotic shape on the hardware side, with the time reversal (or inverse filter) technique on the software side. Using a two step procedure of direct excitation-reception and time reversed excitation-reception, it is possible to focus energy anywhere inside a non-reverberating sample, thanks to the ergodicity of the chaotic cavity. Moreover, the same basic concept can be used to create a virtual phased array using a single channel device. Both experimental data and simulations will be provided to illustrate the concepts. The goal is to use the chaotic cavity transducer concept to enhance the localization of micro-damage coupled to nonlinear elastic wave spectroscopy methods.

9:15

4aPAa4. Generation of sound by flow instabilities in a low-speed helium jet. W. C. Kirkpatrick Alberts, II and David A. Ligon (US Army Res. Lab., ATTN: RDRL-CIE-S, 2800 Powder Mill Rd., Adelphi, MD 20783)

Schlieren images of low-speed helium flow from a nozzle of circular cross section reveal instability in the jet occurring approximately 1 cm from the nozzle orifice. Accompanying this instability is a pronounced whistle. In this presentation, both qualitative and quantitative analyses of the observed phenomenon will be discussed.

9:30

4aPAa5. Parametrically excited oscillations: Where does the energy come from? Murray Strasberg (David Taylor Model Basin, 9500 MacArthur Blvd., West Bethesda, MD 20817-5000, murray.strasberg@navy.mil)

Recent discussions of parametrically excited oscillations seem to have overlooked an obvious question. How is energy transferred to a parametrically excited system to replace the energy lost to dissipative elements? The conventional analysis converts the linear differential equation representing the parametrically excited system into a Mathieu equation which is a lnear differential equation with a periodically varying coefficient, such as a periodically varying stiffness or length. This does not explain the physical mechanism involved in the energy transfer. To indicate how the energy is transferred, several systems are examined here, viz., (1) the simple pendulum where work is done when the length of the pendulum is varied periodically at half the period of the pendulum to increase the amplitude of its oscillation, (2) the taught string where work is done while sinusoidally varying the tension so as to excite transverse vibration at half the frequency of the varying tension, and (3) shape oscillations of gas bubbles in water excited by spherical volume pulsation at twice the frequency of the shape oscillation. These and other examples lead to the suggestion that more is required to excite parametric oscillation than being a solution of a Mathieu equation.

THURSDAY MORNING, 22 APRIL 2010 GRAND BALLROOM III/IV, 10:00 A.M. TO 12:00 NOON

Session 4aPAb

Physical Acoustics: Shock Wave and High Strain Rate Probes of Materials

Albert Migliori, Chair
Los Alamos National Lab., MS E536, Los Alamos, NM 87545

Chair's Introduction—10:00

Invited Papers

10:05

4aPAb1. Materials science at extreme conditions. James Belak (Condensed Matter and Mater. Div., Lawrence Livermore Natl. Lab., P.O. Box 808, L-45, Livermore, CA 94550, belak@llnl.gov)

The response of a material to an extreme stress transient is sensitive to the loading path and the rate of loading. An extreme example is following a phase transformation, during which the microscopic structure of the material completely rebuilds itself. This microstructure, to a large degree, determines the further response of the material. Traditionally, microstructures have been measured by slicing up the pieces and looking inside with methods such as scanning electron micrsocopy and transmission electron micrsocopy. However, at extreme conditions, such as following a phase transformation, one cannot pick up the pieces and must resort to non-destructive probes such as x-ray scattering. Here, we review recent attempts to determine shocked microstructures using non-destructive diffuse and small-angle x-ray scattering and compare to large-scale molecular dynamics simulations. The talk will anticipate future experiments at emerging light sources such as the Linac Coherent Light Source. [This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.]

10:35—10:45 Break

10:45

4aPAb2. High-pressure magnetically driven compression waves in condensed matter. Jean-Paul Davis (Sandia Natl. Labs., P.O. Box 5800, MS-1195, Albuquerque, NM 87185, jpdavis@sandia.gov)

The *Z* machine is a fast pulsed-power machine at Sandia National Laboratories designed to deliver a 100-ns rise-time, 26-MA pulse of electrical current to *Z*-pinch experiments for research in radiation effects and inertial confinement fusion. Since 1999, *Z* has also been used as a current source for magnetically driven, high-pressure, high-strain-rate experiments in condensed matter. In this mode, *Z* produces simultaneous planar ramp-wave loading, with rise times in the range of 300–800 ns and peak longitudinal stress in the range of 4–400 GPa, of multiple macroscopic material samples. Control of the current-pulse shape enables shockless propagation of these ramp waves through samples 1–2 mm thick to measure quasi-isentropic compression response, as well as shockless acceleration of copper flyer plates to at least 28 km/s for impact experiments to measure ultra-high-pressure (~3000 GPa) shock compression response. This presentation will give background on the relevant physics, describe the experimental technique, and show recent results from both types of experiments. [Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract No. DE-AC04-94AL85000.]

11:15

4aPAb3. The influence of shock loading on material properties and dynamic damage evolution. Ellen K. Cerreta, George T. Gray, III, Darcie D. Koller, Curt A. Bronkhorst, Carl P. Trujillo, and Benjamin L. Hansen (T-3, Los Alamos Natl. Lab., MS B216, Los Alamos, NM 87545)

Many commercial and defense applications require structural metals for extreme environments. Specifically, automotive, aerospace, and infrastructure applications need materials with damage tolerance during dynamic loading. To this end, many studies have examined dynamic deformation and damage evolution. These studies have shown that kinetics of loading are critically important to damage evolution of bulk metals. Particularly, in dynamic loading environments in which a shock wave is imparted to the metal, kinetic and spatial effects based on shock wave shape play important roles in damage. These studies also show that depending on crystal structure, shock loading can alter the subsequent properties of a material significantly. However, while these phenomena are gaining acceptance in the dynamic damage community, the ability to predict these phenomena is limited. Here, the influence of dynamic loading across strain rates 103–106/s will be discussed. The role of test platforms and crystallography to examine the influence of kinetics will be tied to changes observed in deformation and damage evolution. It can be shown that isolating the influence of spatial and kinetic effects during dynamic loading is critical to understanding dynamic damage evolution and with this understanding, capabilities for predicting dynamic damage evolution can be advanced.

Contributed Paper

11:45

4aPAb4. Acoustical shock formation in highly nonlinear fluids. Cristian Pantea and Dipen N. Sinha (Mater. Phys. and Applications, MPA-11, Los Alamos Natl. Lab., Los Alamos, NM 87545)

Weak shock formation is investigated for plane sound waves of finite amplitude in fluids with high-acoustical nonlinearity. The shock formation length is related to the sound speed of the fluid c and its nonlinear parameter B/A. The use of fluids with low-sound speed and high parameter of nonlinearity has the advantage that the shock formation can be achieved at much lower pressures. Also, the experiments can be done on a smaller scale because the shock formation length is relatively small. The experiments are performed using short pulsed sound beams produced by a planar transducer with a resonance frequency of 0.5–1 MHz. The propagation medium consists of either different types of fluorocarbon or methanol. Direct pressure measurements of the acoustic waves in the fluid were obtained using a high-frequency calibrated hydrophone. For comparison, a relatively smaller acoustical nonlinear material like water is also investigated. Experimental results related to second and higher harmonics generated in the fluid and their evolution in time along the propagation axis are compared with theoretical time-domain predictions of the Khokhlov–Zabolotskaya–Kuznetsov equation.

Session 4aPP

Psychological and Physiological Acoustics and Musical Acoustics: Music Processing: Neural Mechanisms and Hearing Impairment

Xiaoqin Wang, Chair
Johns Hopkins Univ., Dept. of Biomedical Engineering, 720 Rutland Ave., Traylor 410, Baltimore, MD 21205

Invited Papers

8:00

4aPP1. Individual differences reveal the basis of consonance. Josh H. McDermott (Ctr. for Neural Sci., New York Univ., 4 Washington Pl., New York, NY 10003, jhm@cns.nyu.edu) and Andrew J. Oxenham (Dept. of Psych., Univ. of Minnesota, 75 E. River Rd., Minneapolis, MN 55455)

Some combinations of musical notes are consonant (pleasant), while others are dissonant (unpleasant), a distinction central to music. Explanations of consonance in terms of acoustics, auditory neuroscience, and enculturation have been debated for centuries. These debates have remained largely unresolved, in part, because the various theories are difficult to distinguish with conventional methods. This talk will describe our recent studies applying the method of individual differences to this problem. We measured preferences for musical chords as well as nonmusical sounds that isolated particular acoustic factors, including the beating and harmonic relationships between frequency components. Listeners preferred stimuli without beats and with harmonic spectra, but across over 250 subjects, only the preference for harmonic spectra was consistently correlated with preferences for consonant over dissonant chords. Harmonicity preferences were also correlated with the number of years subjects had spent playing a musical instrument, suggesting that exposure to music amplifies preferences for harmonic frequencies because of their musical importance. Preferences for stimuli lacking beats, in contrast, were not correlated with musical experience. Harmonic frequency relations figure prominently in many aspects of hearing, and our results indicate that they also underlie the perception of consonance. [Work supported by NIH Grant R01DC05216.]

8:20

4aPP2. Music on more than one note: Pitch perception and neural coding of concurrent harmonic tones. Christophe Micheyl and Andrew, J. Oxenham (Dept. of Psych., Univ. of Minnesota, 75 East River Rd., Minneapolis, MN 55455-0344, cmicheyl@umn.edu)

The basis of Western music lies in the combination of simultaneous (harmony) and sequential (melody) harmonic complex tones or notes. Listeners must "hear out" simultaneous pitches and "track" pitch sequences over time. Surprisingly few psychoacoustic studies have studied pitch perception under such "natural" conditions of more than one note at a time. Similarly, the neural mechanisms that support these abilities remain poorly understood. Here, we will review psychoacoustical and neurophysiological findings, which concur to suggest an important role for frequency selectivity in the ability of the auditory system to segregate concurrent notes. In particular, the ability to accurately discriminate changes in the fundamental frequency or, subjectively, pitch of a "target" harmonic complex tone in the presence of a "masker" tone occupying the same spectral region seems to covary with the limits peripheral frequency selectivity. Given that harmonic complex tones are an important class of sound for both music and speech and that frequency selectivity is usually adversely affected by cochlear damage, such results may have important implications for the current understanding of auditory scene analysis of concurrent music and speech sounds and of the listening difficulties experienced by hearing-impaired individuals. [Work supported by NIH Grant R01 DC05216.]

8:40

4aPP3. Auditory and tactile integration in music meter perception. Juan Huang (Dept. of Neurosci. and Zanvyl Krieger Mind/Brain Inst., The Johns Hopkins Univ., Baltimore, MD 21218 and Dept. of Intelligence Sci., Key Lab. of Machine Percept., Peking Univ., Beijing 100871, China), Darik Gamble, Xiaoqin Wang, and Steven Hsiao (Dept. of Neurosci., Dept. of Biomedical Eng., and Zanvyl Krieger Mind/Brain Inst., The Johns Hopkins Univ., Baltimore, MD 21205)

Meter is a fundamental temporal structure of music. The perception of meter is typically inferred from the occurrence of accents in the music surface. Under normal listening conditions, listening to or playing music is usually accompanied by vibro-tactile input which we hypothesize contributes to meter perception. Previous studies have shown that beat perception can occur via purely tactile stimulation. Whether vibro-tactile stimulation can give rise to meter perception and how it interacts with auditory meter perception is unknown. Here we used accent occurrence and strength as cues to study meter perception in subjects performing auditory only, tactile only, and combined auditory-tactile psychophysical discrimination tasks. We find that subjects can perceive meter through purely auditory only and tactile only stimulation. Furthermore, when stimuli were ambiguous, tactile stimulation was found to enhance meter perception when subjects were given weak auditory amplitude-accented cues. Similarly, auditory stimulation enhanced meter perception when subjects were given weak tactile amplitude-accented cues. These results indicate that meter perception is processed cross-modally, and that auditory-tactile integration plays an important role in the neural representation of temporal structures in music.

9:00

4aPP4. Imaging and inducing disorders in musical perception and production. Psyche Loui (Dept. of Neurology, Beth Israel Deaconess Medical Ctr. & Harvard Med. School, 330 Brookline Ave., Palmer 127, Boston MA)

To perceive and produce music accurately, the brain must represent, categorize, plan, and execute pitched information in response to environmental stimuli. Convergent methods from psychophysics, neuroimaging, and noninvasive brain-stimulation with normal and tone-deaf (TD) subjects were employed to show that neural networks controlling pitch perception and production systems include bilateral frontotemporal networks. First, psychophysical data showed that the perception and production of pitch are uncorrelated in TD subjects, suggesting a disconnection between perception and production brain regions. This disconnection was extended in a diffusion tensor imaging study in TD and control subjects: tractography revealed that the arcuate fasciculus, which connects temporal and frontal lobes, is reduced in TD subjects, especially in its superior division in the right hemisphere. This disconnection highlights the importance of frontotemporal interactions in music processing. Finally, to reverse-engineer the perception-production network, transcranial direct current stimulation was applied over superior temporal and inferior frontal regions. Results showed diminished accuracy in pitch matching after stimulation compared to sham control. Taken together, results demonstrate that intact function and connectivity of a distributed cortical network, centered around bilateral superior temporal and inferior frontal regions, are required for efficient interactions with sounds in the environment. [Supported by NIDCD.]

9:20

4aPP5. Neural substrates of spontaneous musical improvisation. Charles J. Limb (Dept. of Otolaryngol.-Head and Neck Surgery, Johns Hopkins Hospital, 601 N. Caroline St., Baltimore, MD 21287)

To investigate the neural substrates that underlie spontaneous musical performance, functional MRI was used to study improvisation in professional jazz pianists. The purpose of the study was to identify the neural substrates that give rise to spontaneous musical creativity, defined as the immediate, on-line improvisation of novel melodic, harmonic, and rhythmic musical elements within a relevant musical context. It was hypothesized that spontaneous musical improvisation would be associated with discrete changes in prefrontal activity that provide a biological substrate for actions that are characterized by creative self-expression in the absence of conscious self-monitoring. By employing two paradigms that differed widely in musical complexity, it was found that improvisation (compared to production of over-learned musical sequences) was consistently characterized by a dissociated pattern of activity in the prefrontal cortex: extensive deactivation of dorsolateral prefrontal and lateral orbital regions with focal activation of the medial prefrontal (frontal polar) cortex. Such a pattern may reflect a combination of psychological processes required for spontaneous creative behaviors such as improvisation, in which internally motivated, stimulus-independent behaviors unfold in the absence of central processes that typically mediate self-monitoring.

9:40

4aPP6. Musical experience impacts hearing speech in noise. Nina Kraus (Auditory Neurosci. Lab., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, www.brainvolts.northwestern.edu)

Musical experience profoundly impacts how sound is transcribed by the nervous system. This influence is likely mediated by cognitive processes such as attention and memory through the corticofugal system [Tzounopoulos and Kraus, Neuron **62**, 463–469 (2009)]. Hearing in noise is difficult for everyone but especially for children with developmental dyslexia and older adults. We have identified objective neural signatures—from the human auditory brainstem—that reflect hearing in noise [Chandrasekaran *et al.*, Neuron **64**, 311–319 (2009) and Hornickel *et al.*, Proc. Natl. Acad. Sci. U.S.A. **31**, 13027 (2009)]. Musicians develop the ability to hear relevant signals embedded in a network of melodies and harmonies. This ability transfers to hearing a target speaker's voice in background noise. We are beginning to understand the biological basis for this perceptual advantage [Parbery-Clark *et al.*, J. Neurosci **29**, 14100–14107 (2009)]. Sensory processing of speech and music is tightly coupled with the cognitive abilities that underlie language and musical expertise; this knowledge can be used to advantage in the consideration of educational and remediation strategies. [Work supported by NSF SGER 0842376.]

10:00—10:15 Break

10:15

4aPP7. Measuring and predicting the quality of nonlinearly distorted music and speech as perceived by hearing-impaired people. Chin-Tuan Tan (Dept. of Otolaryngol., School of Medicine, New York Univ., 550 First Ave., NBV 5E5, New York, NY 10016), Brian C. J. Moore (Univ. of Cambridge, Cambridge CB2 3EB, United Kingdom), and Mario Svirsky (School of Medicine, New York Univ., New York, NY 10016)

The goals of this study were to characterize and model the perception of nonlinearly distorted speech and music by hearing-impaired listeners. Hearing-impaired listeners were asked to rate the perceived quality of speech and music that had been subjected to various forms of nonlinear distortion, some of which are inherent to certain hearing aid designs including (1) hard and soft, symmetrical and asymmetrical clipping; (2) center clipping; (3) "full-range" distortion, produced by raising the absolute magnitude of the instantaneous amplitude of the signal to a power (8800;1), while preserving the signal of the amplitude; (4) automatic gain control (AGC); (5) output limiting. Stimuli were subjected to frequency-dependent amplification as prescribed by the "Cambridge formula" before presentation via Sennheiser HD580 earphones. The pattern of the rating was reasonably consistent across subjects with only two of ten subjects not making consistent ratings. The mean ratings were not lower with increasing amount of soft or center clipping or when the compression ratios of the AGC and output limiting were increased. The deleterious effects produced by these nonlinear distortions may have been offset by the beneficial effects of improving audibility and compensating for loudness recruitment. [Work supported by Deafness Research Foundation.]

10:35

4aPP8. Music through hearing aids: Perception and modeling. Kelly Fitz and Martin F. McKinney (Starkey Labs., 6600 Washington Ave. S, Eden Prairie, MN 55344, firstname_lastname@starkey.com)

Historically, the primary focus of hearing aid development has been on improving speech perception for those with hearing loss. Modern-day hearing-aid wearers, however, face many different types of acoustic signals, such as music, that require different types of processing. Music signals differ from speech signals in a variety of fundamental ways, and relevant perceptual information is conveyed via different signal attributes in the two types of signals. The research described here is an effort to improve music perception in listeners with hearing impairment. First, methods have been developed to quantitatively measure deficits in music perception for impaired and aided listeners. Second, specific perceptual features have been evaluated as to their relative importance in the successful perception of music and that information has been used to guide signal processing development. Finally, the relevant perceptual features have been modeled, and the models have been used to evaluate and compare signal processing algorithms designed to improve music perception through hearing aids. An overview of our research will be presented along with key recent results.

Contributed Papers

10:55

4aPP9. A novel physiological mechanism for pitch encoding in monkey primary auditory cortex. Yonatan I. Fishman (Dept. of Neurology, Albert Einstein Coll. of Med., 1300 Morris Park Ave., Bronx, NY 10461, yonatan.fishman@einstein.yu.edu) and Mitchell Steinschneider (Albert Einstein Coll. of Med., Bronx, NY 10461)

Pitch is a fundamental perceptual attribute of hearing. While auditory cortex is implicated in pitch perception, how pitch is represented at the cortical level remains unclear. The present study examines a novel hypothesis for how the pitch of pure tones and of harmonic complex tones, with or without the fundamental frequency, is encoded in primary auditory cortex (A1): pitch is represented non-topographically by the temporal distribution of population activity in A1. Sounds of lower pitch evoke a greater proportion of sustained multiunit activity (MUA), relative to initial onset MUA, than sounds of higher pitch, such that the temporal distribution of MUA systematically varies with pitch. Pure tones and harmonic complexes with the same pitch evoke a similar proportion of sustained MUA. The temporal distribution of MUA is largely invariant to changes in stimulus parameters (e .g., level and relative phase) that leave the perceived pitch unchanged. Timing of AEP components recorded in superficial layers of A1 parallels similar pitch-related changes in AEPs recorded in humans. Coding of perceptual qualities based on the time course of neural activity has been proposed in other sensory modalities (e.g., olfaction) and offers a novel alternative to topographic representations of pitch at the cortical level.

11:10

4aPP10. Neural correlates of sensory consonance and dissonance in primate primary auditory cortex. Mitchell Steinschneider (Dept. of Neurosci., Albert Einstein Coll. of Med., 1300 Morris Park Ave., Bronx, NY 10461, mitchell.steinschneider@einstein.yu.edu) and Yonatan I. Fishman (Albert Einstein Coll. of Med., Bronx, NY 10461)

The perception of consonance and dissonance of isolated chords (sensory consonance/dissonance) is fundamental to music appreciation. Consonant chords are composed of tones related to each other by simple frequency ratios (e.g., perfect fifth 3:2), whereas dissonant chords are composed of tones related to each other by complex ratios (e.g., minor second 256:243). Dissonance is thought to be due to the perception of beats (modulation frequencies $<$ 20 Hz) or roughness (modulation frequencies from 20–250 Hz), which occur when two or more components of a complex sound are separated from one another in frequency by less than the width of an auditory filter (i.e., critical bandwidth). These unresolved frequency components interact in the auditory periphery to produce amplitude-modulated (AM) fluctuations in the composite waveform envelope. We demonstrate that the magnitude of neuronal phase-locking to these AM fluctuations in primary auditory cortex of awake monkeys correlates with the perceived consonance/dissonance of musical chords and parallels human perception of roughness. This correlation is displayed by population activity, as measured by auditory evoked potentials (AEPs), current source density, and multiunit activity. We further demonstrate that phase-locking of AEPs in Heschl's gyrus of humans is similar to that seen in the monkey.

11:25

4aPP11. A biomimetic multi-resolution spectrotemporal model for musical timbre recognition. Kailash Patil and Mounya Elhilali (Dept. of Elec. & Comput. Eng., 3400 N. Charles St., Barton 105, Baltimore, MD 21218)

Timbre is usually defined as that property of sound that is neither pitch nor loudness that helps in identifying sounds. Such a definition of *what it is not* is in itself problematic and unsatisfactory. The complexity of characterizing a timbre space stems from the intricate interactions between spectrotemporal dynamics of sound, which overcasts the simple description of individual acoustic dimensions as usually captured by techniques such as multidimensional scaling. By contrast, sound encoding in the mammalian auditory system, and particularly its sensitivity to spectrotemporal modulations, offers a rich feature space to explore perceptual representations of timbre. In the present work, the problem of musical timbre modeling was casted as a generative model based on a reduced set of multi-resolution spectrotemporal features inspired from encoding of complex sound in the auditory cortex. A probabilistic system based on Gaussian mixtures was built to perform a timbre recognition task on a data set consisting of 12 instruments. The system was able to achieve an accuracy of 98%, hence corroborating the claim that physiologically inspired features are effective in properly outlining a timbre space. These results could have significant implications in terms of better understanding the neural correlates of timbre perception.

Session 4aSA

Structural Acoustics and Vibration: Applications of Structural Acoustics and Vibration I

Robert M. Koch, Chair
Naval Undersea Warfare Center, 1176 Howell St., Newport, RI 02841-1708

Contributed Papers

9:00

4aSA1. Comparison of techniques of acoustically interrogating a fluid-filled pipe. Curtis F. Osterhoudt, Christopher Dudley, and Dipen N. Sinha (MPA-11, Los Alamos Natl. Lab., Los Alamos, NM 87545, cfo@lanl.gov)

Through-transmission of sound in a fluid-filled pipe was experimentally investigated, with comparison to a one dimensional computational model. Pipe curvature is shown to be a contributor to deviations from the model, as is dynamical interaction between the finitely yielding pipe walls and the fluid contents. Various techniques of acoustically interrogating the system are considered. These include the different information which may be extracted via continuous-wave acoustical excitation, swept-frequency interferometry, and tone- and chirp-burst excitations. Each of these techniques has their advantages and disadvantages, especially in situations where short-time measurements must be made, and a version of the gain-bandwidth product must be taken into account. Some algorithms for automatically extracting fluid sound-speed are discussed, and it is shown that no one dominates under all circumstances. [This work was supported by Chevron USA.]

9:15

4aSA2. Sonar-induced strains and stresses in an absorbing solid elastic sphere. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA, 02543), David T. I. Francis (Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom), Mario Zampolli, and Tim van Zon (TNO Defense, Security and Safety, 2509 JG The Hague, The Netherlands)

The interaction of harmonic plane and spherical pressure waves with an absorbing solid elastic sphere is modeled. The analytic, boundary-element-method (BEM) and finite-element-method (FEM) solutions for the internal displacement, strain, and stress fields are evaluated numerically for 50-mm-diameter solid spheres of poly(methyl methacrylate) (PMMA) and high-density polyethylene (HDPE) in an immersion medium of mass density 1000 kg/m^3 and sound speed 1500 m/s. The mass density and longitudinal and transverse sound speeds of the PMMA sphere are 1191 kg/m^3, 2690 m/s, and 1340 m/s, respectively. The corresponding properties of the HDPE sphere are 957 kg/m^3, 2430 m/s, and 950 m/s. The two materials are assumed to have a hysteresis-type absorption, hence with constant product of absorption coefficient and wavelength. The respective values of this product for longitudinal and shear waves for PMMA are assumed to be 0.19 and 0.29 dB, and for HDPE, 0.40 and 1.20 dB. Each of two frequencies is considered, 10 and 100 kHz, for which the wavenumber-radius product is $\pi/3$ and $10\pi/3$, respectively. Results for the three solution methods are compared. [Work partly supported by NOPP through ONR Award No. N000140710992.]

9:30

4aSA3. On the cancellation of acoustic waves scattered from an elastic sphere. Matthew D. Guild (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, mdguild@arlut.utexas.edu), Andrea Alù, and Michael R. Haberman (The Univ. of Texas at Austin, Austin, TX 78713-8029)

Recent research has suggested the possibility of creating non-absorptive elastic covers that eliminate the acoustic field scattered from an elastic object, also known as acoustic cloaks [Norris, Proc. R. Soc. London, Ser. A **464**, 2411 (2008)]. The work presented here employs the scattering cancellation technique [Alù and Engheta, Phys. Rev. E **72**, 016623 (2005)] to investigate the effectiveness of a single isotropic elastic layer to cloak an elastic sphere. The presentation discusses the benchmarked analytical and finite element scattering models which were employed to explore the design space of the cloaking layer. Parametric studies showing the influence of cloak stiffness and geometry on the frequency dependent scattering cross section are then presented. These case studies clearly illustrate the fundamental physical behavior leading to the observed reduction in scattering cross section at design frequencies. Finally, material selection and the creation of composite materials required to produce a functional scattering cancellation layer are presented and discussed.

9:45

4aSA4. Effective mass density of a random configuration of cylinders. Francine Luppé (LOA-GOA, FRE CNRS 3102, Univ. of Le Havre, pl.R. Schuman, 76610 Le Havre, France, francine.luppe@univ-lehavre.fr), Jean-Marc Conoir (Univ. Paris 6, 75005 Paris, France), and Pascal Pareige (Univ. of Le Havre, 76610 Le Havre, France)

The dynamic effective mass density of a random distribution of $n0$ cylinders/m^2 in an ideal fluid is looked for. The Fikioris and Waterman approach is used to obtain the reflection coefficient of the random half-space at the plane interface with the ideal fluid. This coefficient is expanded into powers of $n0$, using Linton and Martin's expansion of the wavenumber of the coherent wave. The reflection coefficient is then compared to that obtained when a homogeneous viscous fluid replaces the random medium. When the two reflection coefficients are equal, the random fluid is acoustically equivalent to the viscous one, which is thus considered as the effective fluid. The coherent wave in the random medium is described as the acoustic mode in the effective fluid, with the shear viscosity of the latter being set to zero. Equating the two reflection coefficients provides an effective mass density that depends on frequency and on the incidence angle, except at low frequency. The angle dependence is discussed.

10:00

4aSA5. Coherent backscattering enhancement in cavities. Simple shape cavity revisited. Stefan Catheline, Thomas Gallot, Philippe Roux, Guillemette Ribay, and Julien de Rosny (Laboratoire de Gophysique Interne et Tectonophysique (LGIT), CNRS Universit de Grenoble, France, stefan.catheline@ujf-grenoble.fr)

Coherent backscattering effect (CBE) is classically introduced in disordered, random, or chaotic media. In this work, the attention is focused on simple parallepipedic cavities since, contrarily to a widespread idea, CBE can also be observed for a pure-tone source in a one-dimensional (1-D) cavity. This approach is of two-fold interest. First, analytical computations predict a dimensional dependence of the coherent backscattering enhancement according to a $^{(3/2)}d$ law, d being the dimensionality of the cavity, that have not yet been compared to experiments. Second, it opens a new ballistic interpretations for which each multiply reverberated path is associated with more (rectangle and parallelepipedic cavities) or less (1D cavity) than one single reciprocal counterpart. This paper is the first of two, the second paper dealing with some impacts of symmetry on CBE.

10:15—10:30 Break

10:30

4aSA6. Coherent backscattering enhancement in cavities. Highlight of the role of symmetry. Thomas Gallot, Stefan Catheline, and Philippe Roux (Laboratoire de Gophysique Interne et Tectonophysique (LGIT), CNRS & Universit de Grenoble, France, thomas.gallot@obs.ujf-grenoble.fr)

Through experiments and simulations, some consequences of symmetry on coherent backscattering enhancement (CBE) in cavities are reported here. First, CBE outside the source is observed (a) on a single symmetric point in a one dimensional cavity, in a tw0-dimensional (2-D) circular membrane, and in a 2-D symmetric chaotic cavity; (b) on three points in a rectangle; and (c) seven in a parallelepiped. Second, existence of enhanced intensity lines and plans in 2-D or three-dimensional cavities is demonstrated. Third, we show how antisymmetry is responsible for the existence of a coherent backscattering decrement with a dimensional dependence of (12)d law, *d* being the dimensionality of the cavity.

10:45

4aSA7. How does the scattering from an empty cylinder change with the level of filling? Duncan P. Williams (Dstl Physical Sci., Porton Down, Salisbury SP4 0JQ, United Kingdom), Richard V. Craster (Univ. of Alberta, Edmonton T6R 2Y9, Canada), and Samuel D. M. Adams (Imperial College, London SW7 2BZ, United Kingdom)

The acoustic scattering from elastic objects features in many tasks stretching from the use of sonar to search for underwater objects such as submarines and mines to the inspection of shipping containers and cargo screening. The properties of canonical objects, such as cylindrical and spherical shells, which are completely filled with fluid, or immersed in fluid, have been widely studied, but typically stop short of studying shells containing different levels of filling. Not knowing the extent of the effect of filling on the scattering from different objects can limit how well one can discriminate between similar objects or inspect the interior of objects and, for example, find contraband or other suspicious materials. This paper looks at the two-dimensional response of a partially filled elastic cylinder. A computational method to model non-uniform domains is introduced and the use of perfectly matched layers is discussed. Results are shown for combinations of cylinders and different levels of filling. In particular, we show how the response of the cylinder depends on the level of the filling. The results are used to comment on the usefulness, or otherwise, of the response to estimate the nature and level of filling based on short- or long-range observations.

11:00

4aSA8. Modeling of a two-transducer through-wall ultrasonic communication system. Sebastian Roa Prada, Kyle R. Wilt, Henry A. Scarton, Gary J. Saulnier, Jonathan D. Ashdown, Pankaj K. Das, and Andrew J. Gavens (Dept. of Mech. Eng., Rensselaer Polytechnic Inst., 110 Eighth St., Troy, NY 12180, sebastian-roa27@hotmail.com)

Ultrasound at 1 MHz is used as a carrier of information across solid walls without penetration. A communication channel is created by bonding two transducers on either side of a solid wall. The outside transducer transmits a continuous wave generating an acoustic field. The inside electronics are powered by harvesting the electrical output of the inside transducer. Digitized bits, representing the value of an analog signals measured on the inside (such as temperature), are used to alternate the electrical load of the inside transducer between two finite values. Changes in the electrical load of the inside transducer modify its acoustical impedance as seen by the incident waves coming from the wall, which in turn modulates the amplitude of the reflected signal. This modulated wave is detected at the electrical terminals of the outside transducer, where it is then demodulated to recover the data. The mechanical components of the system are modeled in connection to the electronic circuits by means of electro-mechanical analogies. Simulation of the communication system is performed using the electric circuit simulation package PSPICE. Simulation, finite element solutions, and experimental results are presented and discussed. Digital data communication rates exceeding 50 000 bits/s are achieved.

11:15

4aSA9. Relationships between structural energy density, power flow, and their influence on acoustic intensity. Jeff M. Fisher, Dan A. Manwill, Jon D. Blotter, Scott D. Sommerfeldt, and Kent L. Gee (Dept. of Mech. Eng., Brigham Young Univ., 435 CTB, Prov, uT jefffish@nm.byu.edu)

Power flow in structures has been a topic of research for the past few decades. Many different methods for determining structural power flow magnitude and direction have been developed and proven. Structural energy density methods have likewise been developed. In this paper, a study of both structural energy density and power flow in a plate and their influence on the acoustic field is presented. Structural energy density and power flow were determined using equations typically used with accelerometer arrays but adapted for use with a scanning laser doppler vibrometer. Experiments were performed from 25 to 100 Hz by exciting the plate at single frequencies but using different excitation points and methods at each frequency. In some cases two sources of excitation were used to alter the power flow response, giving a broader basis for concluding relationships. Both speaker and shakers were used as sources. The acoustic intensity in the room was calculated at approximately 3 in. from the vibrating plate. Relationships between structural energy density, power flow, and the acoustic field variables are presented.

11:30

4aSA10. Real-time active control of structural energy density and structural power flow. Daniel A. Manwill (dmanwill@byu.net), Jeff M. Fisher (Dept. of Mech. Eng., Brigham Young Univ., 435 CTB, Provo, UT 84602), Scott D. Sommerfeldt, Kent L. Gee, and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Structural energy density and structural power flow have long been used as metrics in the active control of vibrating structures. The greater portion of this previous work has focused on frequency-domain methods which incorporate assumptions about the relative contributions of near-field and far-field energy components. This paper describes the implementation of filtered-x-based time-domain control schemes which utilize 9- and 13-accelerometer arrays to estimate and control structural energy density and structural power flow, respectively. Experiments were performed on a clamped steel plate excited and controlled by various combinations of loudspeakers and electrodynamic shakers in a frequency range from 25 to 100 Hz. Analog circuitry was used to estimate spatial derivatives and reduce channel count. The development of control laws incorporating the effects of the analog circuitry is presented. Control attentuation results are given, sensor placement is discussed, and implementation challenges are addressed. [This work is supported by NSF Grant 0826554.]

Session 4aSC

Speech Communication: Speech Perception and Perceptual Adjustment (Poster Session)

Cynthia Clopper, Chair
Ohio State Univ., Dept. of Linguistics, 1712 Neil Ave., Columbus, OH 43210

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors or odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors or even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

4aSC1. Perceptual and production training of allophones and phonemes in Spanish. Wendy Herd (Dept. Linguist., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044, wenherd@ku.edu)

American English learners of Spanish often do not acquire native-like pronunciation of intervocalic /d/, tap, and trill in words like CODO "elbow," CORO "choir," and CORRO "I run." The trill proves difficult because it does not exist in English. Although the tap exists as an allophone of /t/ and /d/ in English, students of Spanish must learn to process it as a phoneme rather than an allophone. Similarly, learners have difficulty acquiring the spirantization of voiced stops, where intervocalic /d/ is produced as a voiced dental fricative or approximant. This study investigates whether American English learners of Spanish can be trained to perceive and produce the intervocalic /d/, tap, and trill contrasts in Spanish. Both perceptual and production training methods were used. Past research has reported that perceptual training alone improves both perception and production, and that production training alone improves both as well; however, the production training studies have not been limited to production as trainees have been able to listen to the training stimuli. This study systematically controls both training modalities and introduces a third training methodology that includes both perception and production to discover whether perceptual, production, or combination training is most effective. [Research supported by NSF.]

4aSC2. Production and perception of Taiwan Mandarin syllable contraction. Grace Chen-Hsiu Kuo (Dept. of Linguist., Phonet. Lab., UCLA, Los Angeles CA 90095, gracekuo@humnet.ucla.edu)

Taiwan Mandarin syllable contraction is an optional lenition process which involves the elision of the intervocalic segments and the merger of the tonal elements of two syllables. Here it is shown that syllable contraction is gradient and non-neutralizing. In a production experiment, 20 subjects read a list of minimal sentence pairs, containing disyllabic contractable words and matched monosyllabic lexical words, at two speech rates (88 and 144 beats/min), three times each. Degree of contraction was measured as the depth of the intensity trough between two syllables [Mermelstein 1975], with a trough depth (TrD) of zero meaning fully contracted. 8% of disyllables were somewhat contracted (TrD between 0 and 2 dB) while 29% were fully contracted (TrD = 0 dB). Fully contracted disyllables were compared on several other acoustic measures to their monosyllabic counterparts and were found to differ most strongly in duration. In a perception experiment, items from the production experiment were presented to 35 listeners for forced-choice identification as disyllabic or monosyllabic words. Accuracy was generally high; reaction times were slower for lexical monosyllables, which were sometimes labeled as contracted. Thus, even when fully contracted syllables were produced, they remained acoustically and perceptually distinct from monosyllables.

4aSC3. Prosodic perception and production of English- and Chinese-native speakers. Amanda Rodriguez and Chang Liu (Dept. of Commun. Sci. and Disorder, The Univ. of Texas at Austin, 1 University Station A1100, Austin, TX 78712)

Our previous studies have shown that English-native adult speakers demonstrated categorical perception in tonal identification of speech and nonspeech sounds as typically as Chinese-native adult speakers. The purpose of this study was to investigate whether prosodic perception of English speech sounds was different between English- and Chinese-native listeners. F0 contour was manipulated from falling to rising patterns for the target word embedded in a short sentence. Listener's task was to identify the prosody of each sentence, either question or statement. Preliminary results suggested that both groups of listeners showed typical categorical perception, while the two groups had significant difference in categorical boundary. The difference in categorical boundary for prosodic perception between the two groups of listeners was similar to our previous findings in tonal perception, likely due to the difference in the listener's language background. English sentences with statement and question will be recorded for the two groups of listeners. The relationship between prosodic perception and perception will be discussed. [Work supported by The University of Texas at Austin, Undergraduate Research Fellowship.]

4aSC4. Influence of perceptual training of syllable codas for English consonants on sentences. Teresa Lopez-Soto (Dept. English Lang., Univ. of Seville, c/ Palos de la Frontera, s/n 41004 Sevilla, Spain, teresals@us.es) and Diane Kewley-Port (Indiana Univ., Bloomington, IN 47405)

This study extends our previous report suggesting that moderate amounts of speech perception training were associated with improved speech production. This study also examined perception in sentences. A new group of 8 Spanish speakers (<10 year residence in US) trained on 13 English final consonants in syllables using SPATS software. On days 1 and 7 the group participated in perception and production tasks with the 13 codas (pre- and post-tests). Perception tests included coda identification in both syllable stimuli and in 18 sentences extracted from the IEEE corpus (90 keywords). Results showed the following. (1)With 5-h training of the 13 codas in syllable stimuli, perception improved significantly (>10%) in the syllable post-test, with smaller improvement in the sentence post-test (>7%). (2) Although average improvement was small for codas tested in sentences, several improved substantially (>20%), while a few decreased considerably after perception training. (3) Absolute values in both post-tests show similar performance, near 90%, although interesting differences for specific consonants were noted. The results showed that perception of sounds when measured in different linguistic contexts (syllables versus sentences) rendered different results. The study lays ground for investigating how cross-language perception of individual sounds is influenced by the phonetic context.

4aSC5. Cue weighting and variability in perception and production. Jiwon Hwang (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376)

The Korean single liquid phoneme shows an allophonic variation: lateral [l] occurs in coda and tap [r] in onset. Intervocalically, they may appear contrastive as tap or geminate lateral [ll] ([iri] wolf vs [illi] reason), differing in duration and laterality. Kim [(2007)] demonstrated that Korean listeners identified a shortened geminate lateral as geminate /ll/ rather than tap, despite the fact that the duration of edited stimuli was matched for tap. The current study examines whether the weighting of laterality cues over duration cues for [ll] vs [r] is motivated by the native language acoustics. Korean speakers produced 24 Korean words [(C)V_V], containing [ll] or [r] in two

speech modes (in a carrier sentence vs in isolation). The duration of [ll] was significantly longer than tap, but it varied greatly depending on the speech mode while tap did not. The intensity of tap was significantly lower than [ll] generally, but it showed a greater variability. However, F3 at the offset of the preceding vowel was lower for tap than for [ll] regardless of the speech mode. The consistent use of F3 in production supports the perceptual cue weighting pattern where Korean listeners rely more on spectral cues than duration for intervocalic [ll]-[r] contrast.

4aSC6. The production and perception of English consonant sequences by Japanese-speaking learners of English. Mieko Sperbeck and Winifred Strange (Dept. of Linguist., City Univ. of New York, the Graduate Ctr., 365 Fifth Ave., New York, NY 10016)

This study reports the second part of a study investigating vowel insertion phenomena among Japanese speakers. A previous study [Sperbeck (2009)] that measured categorial discrimination demonstrated that some contrastive CCV versus CəCV sequences were hard for Japanese listeners (72% correct overall). The current study explored difficulties in production and how production correlated with perception. Nonsense words were constructed as the stimuli. They were of the form/CC(C)ani/ and /CəC(C)ani/, where CC(C) combinations were /sp, sk, pl, kl, bl, gl, spl, skl/. A delayed imitation task was used to assess production. Participants heard a native speaker's productions (e.g., Say blani now) twice, produced the target word in isolation (e.g., blani), and then produced it in the carrier sentence (e.g., I said blani now). The latter was scored in this study. Two phonetically trained native English speakers perceptually transcribed the productions. Results showed that the overall percent correct was 66% (SE = 3.17) among Japanese speakers. There was a significant correlation between perception and production performance (rho = + 0.715. $p<0.01$). However, the major error type was vowel deletion, rather than vowel epenthesis in producing the CəCV tokens. The relationship between perception and production among L2 learners will be discussed.

4aSC7. The discrimination, perception, and production of two German /r/ allophones by two groups of American English speakers. Dilara Tepeli (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706)

The German /r/ sound is one of the most difficult sounds for American English (AE) speakers learning German as a foreign language. Part of this difficulty may be due to its rich phonetic variation. The standard German /r/ variant [R] and dialectal variant [R′] are achieved by varying the tongue constriction degree while keeping place of articulation constant [Schiller and Mooshammer (1995)]. The close articulatory proximity of these allophones provides an opportunity for testing the relationship between perception and production in L2 sound acquisition. The aim of this study is to investigate how well experienced AE speakers and naive AE speakers can discriminate and produce the difference between the uvular fricative [R] versus the uvular trill [R′]. Two groups of AE subjects who participated in an imitation study were prompted to produce single words beginning with either [R] or [R′]. Subjects also participated in a discrimination and categorization test. Preliminary results suggest that inexperienced AE can discriminate [R] versus [R′] well. They often perceive the sounds as /h/ and are more successful at producing [R′] than [R]. Experienced speakers also discriminate the two sounds well, perceive both sounds as the German /r/ and struggle more with producing [R′] than [R].

4aSC8. Auditory feedback shifts in one formant cause multi-formant responses. Shira Katseff (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, skatseff@berkeley.edu), John F. Houde (Univ. of California at San Francisco, San Francisco, CA 94143-0444), and Keith Johnson (Univ. of California, Berkeley, Berkeley, CA 94720-2650)

Talkers are known to compensate for experimentally-induced shifts in auditory feedback. In a typical experiment, talkers might hear their F1 feedback shifted (so that [ɛ] sounds like [æ], for example), and compensate by lowering F1 in their subsequent speech. Typically, compensation is assumed to directly oppose the action of the feedback shift and is measured in terms of the shifted parameter. In this study, we instead find that sensitivity to altered auditory feedback is multidimensional: subjects respond to altered F1 feedback by changing their F2 production and vice versa. In particular, subjects whose [i] is heard as [ɪ], a shift primarily in F1, compensated by producing a higher F2, while subjects whose central vowel [ʌ] was heard as [ɛ] or [o], a shift primarily in F2, compensated by producing a higher or lower F1. We argue that it is insufficient to consider auditory sensitivity in terms of a single formant and suggest that this method of altering auditory feedback is a practical tool for investigating the psychological reality of formants and their combinations.

4aSC9. Perception of vocal imitations and identification of the imitated sounds. Guillaume Lemaitre, Arnaud Dessein (IRCAM, 1 place Stravinsky, 75004 Paris, France), Karine Aura (Universit de Toulouse le Mirail, 31058 Toulouse, France), and Patrick Susini (IRCAM, 75004 Paris, France)

We report two studies investigating how vocal imitations enable the recognition of the imitated sounds. First, we asked couples of participants to listen to series of everyday sounds. One of the participants ("the speaker") had then to describe a selected sound to the other one (the "listener"), so that he could "guess" the selected sound. The results showed that, spontaneously, the speakers used, among other para-linguistic cues, large numbers of vocal imitations. Moreover, they suggested that the identification performances were increased when vocal imitations were used, compared to only verbal descriptions. Second, we sampled 28 sounds across an experimental taxonomy of kitchen sounds and required laypersons to vocally imitate these sounds. Another group of participants was then required to categorize these vocal imitations, according to what they thought was imitated. A hierarchical cluster analysis showed that, overall, the categories of vocal imitations fitted well with the categories of imitated sound sources. By using finer analysis techniques, we also showed that some imitations inconsistently clustered. On the other hand, the consistent clusters of imitations were perfectly predicted by a few acoustical descriptors. We therefore conclude that vocal imitations of sounds contain enough information for the recognition of the imitated sounds.

4aSC10. The perception and acoustic features of Korean ditropic sentences. Seung-yun Yang, Ji Sook Ahn, and Diana Van Lancker Sidtis (Dept. of Communicative Sci. & Disord., NYU, 665 Broadway, Ste. 900, New York, NY 10003)

Ditropic sentences are utterances that convey either a literal or an idiomatic meaning (e.g., It broke the ice). This study investigated listener's ability to discriminate between literal or idiomatic meanings and examined the acoustic features contributing to this distinction. Ten ditropically ambiguous Korean sentences were audio-recorded by four native speakers of Korean. Each utterance was produced twice with either a literal or idiomatic meaning. Fifteen native Korean subjects listened to a randomized presentation of these utterances singly and in pairs without other context and identified each as literal or idiomatic. Listeners successfully discriminated the intended idiomatic or literal meanings (singletons = 70.65%, pairs = 75 .67%). These results were consistent with those of Van Lancker and Canter [(1981)] for English ditropic sentences. Each utterance was acoustically analyzed in terms of means and variations in fundamental frequency, duration, and intensity. Analyses of variance revealed significantly longer durations and greater variation in syllable duration for literal than idiomatic sentences, whereas idiomatic sentences were characterized by significantly greater variation in intensity than literal sentences. Some prosodic cues for Korean differed from those found previously for English [Van Lancker *et al.* (1981)] and French [Abdelli-Baruh *et al.* (2007)]. These results further understanding of use of prosody in sentential linguistic contrasts.

4aSC11. Discrimination and identification of synthetic [da]-[ga] sounds by adults and children 4–6 years of age. Kelly Richardson (Dept. of Communicative Disord. and Sci., Univ. at Buffalo, 3435 Main St., Buffalo, NY 14214, kcr2@buffalo.edu) and Joan Sussman (Univ. at Buffalo, Buffalo, NY 14214)

Identification of sounds along the alveolar-to-velar contrast has been shown to be different for children born with clefts of the palate compared to other children [e.g., Whitehill *et al.* (2003)]. However, little information concerning discrimination abilities of children, in general, is known. The current study used a seven-step continuum of synthetic consonant-vowel syllables changing from "da" to "ga" by the 40 ms third formant frequency transitions [Sensimetrics Corporation (1995)]. Each listener heard 180 trials of a "change no-change" discrimination paradigm with two, three, four, and

six-step comparisons to the endpoint, stimulus number 1 "da." Listeners were asked to verbally respond as to whether there was a "change" or "no-change" in the sounds presented. Results revealed that as the stimulus contrast increased, adults became better at discriminating the comparisons, whereas children's performance remained poor for all comparisons. Adults also showed less variability in their responses as the contrasts grew larger. In the identification task, young children displayed a much smaller velar category compared to the adult group. Furthermore, children were significantly poorer than adults at identifying the endpoint stimuli, as well as within-category exemplars. Results are compared to perception of other places of articulation by children and adults.

4aSC12. Misperceptions of foreign language phonology. Julia Yarmolinskaya, Colin Wilson, and Brenda Rapp (Dept. Cognit. Sci., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

For many years it has been assumed that when hearing familiar sounds in unfamiliar combinations, listeners will perceive the sounds accurately. In the recent years, this assumption has been challenged [Halle *et al.* (1998) and Berent *et al.* (2006)]. The present study investigates what listeners actually hear when presented with familiar consonants in unfamiliar combinations. A group of monolingual native English speakers was asked to transcribe Russian words containing only consonants attested in English but presented in a two-, three-, and four-consonant combinations (CC, CCC, and CCCC clusters) which are not legal in English. The results indicate that the accuracy of response and the nature of errors depend on the type of cluster. CC clusters (e.g., /pn/) were most often misperceived as containing a vowel between the two consonants, and perception accuracy was well-explained by phonological principles, such as sonority. On the other hand, CCC (e.g., /vzb/) and CCCC (e.g., /fstr/) cluster transcriptions contained relatively few vowel insertions, but many deletions and substitutions and their accuracy were better explained by acoustic factors, such as voicing of the consonants. These results suggest that speech perception is influenced by both phonological and acoustic factors.

4aSC13. Effects of listener experience with foreign accent on perception of accentedness and speaker age. Paul Rodrigues (Dept. of Linguist., Indiana Univ., 1021 E, Third St., Bloomington, IN 47405, prrodrig @indiana.edu) and Kyoko Nagao (Nemours A.I. duPont Hospital for Children, Wilmington, DE 19803)

The current study examined the effects of foreign accent and listener experience on the perception of a speaker's age and native language. Ten audio stimuli were prepared from the recording of five Arabic speakers and five English speakers (18–79 years old) from the Speech Accent Archive [Weinberger (2009)] for the perception experiment. Thirty native speakers of English participated in the perception experiment through Amazon's Mechanical Turk website, estimated the speaker age, rated the speaker's accentedness, and estimated the native language of the speaker. The listeners were divided into two groups based on their experience with foreign accented English (experienced and inexperienced groups). Higher correlation was found between perceived age (PA) and actual chronological age (CA) for the native (English) stimuli than for the Arabic-accented stimuli in both listener groups. The correlation between PA and CA was higher in the experienced listeners than in the inexperienced listeners. Accentedness rating suggests that the inexperienced listeners tend to rate both native and non-native speakers neutrally on the scale. The results suggest that experiences with any foreign accented speech facilitates identification of age from speech and help to form the ability to perceive differences in the degree of foreign accented speech.

4aSC14. Word segmentation of American English /s/ in semi-spontaneous speech. Dahee Kim (Dept. of Linguist., The Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210-1298, daheekim@ling .osu.edu), Christine Szostak, Colin Widmer, and Mark A. Pitt (The Ohio State Univ., Columbus, OH 43210-1287)

To comprehend spoken language, listeners need to find words from a continuous stream of speech sounds. Little work has explored whether there are reliable acoustic cues to word boundaries in conversational speech, which is highly reduced and under-articulated, potentially creating ambiguities at word boundaries. Segmentation may be even more difficult when the same segment repeats at a word boundary, ending the preceding word and beginning the following word (e.g., gas station). Segmentation in this environment was investigated by examining the production and perception of fricative /s/ in semi-spontaneous speech. Twenty talkers produced sentences containing ambiguous two-word sequences with /s/ between the two words. All sequences are interpretable in three ways (e.g., grow snails, gross snails, and gross nails) depending on how the frication is segmented. Acoustic analyses of the production data examined whether there are acoustic cues distinguishing the three versions of the ambiguous sequences. Listening experiments using the talkers' productions as stimuli evaluated the degree of ambiguity in the tokens and identified acoustic cues that listeners use to segment the two words. Results will be discussed in the context of theories of speech perception and word segmentation.

4aSC15. Perception of prosodic boundaries in spontaneous speech with and without silent pauses. Yoonsook Mo and Jennifer Cole (Dept. of Linguist., Beckman Inst., Univ. of Illinois, Urbana-Champaign, 1420 N. Mathews Ave., Urbana, IL 61801, ymo@illinois.edu)

In speech comprehension, listeners attend to variation in multiple acoustic parameters encoding prosodic structure. Given the multiplicity of acoustic cues, we ask whether prosody perception is dependent on any individual cue or whether acoustic redundancy encoding prosody supports robust prosody perception in the absence of an individual cue. The present paper reports on a study of boundary perception in spontaneous speech with and without silent pause as a boundary cue. Prior studies show that in read speech, silent pause is important for boundary perception, while in spontaneous speech, listeners can detect boundaries without pauses. Our study tests the role of pause in boundary perception with two versions of 36 short speech excerpts from the Buckeye Corpus: one with pauses intact and another with all pauses truncated to 20 ms. In real-time transcription tasks based only on auditory impression, boundary locations were marked by 74 subjects for the intact stimuli and by an additional 15 subjects for truncated excerpts. Inter-transcriber agreement was comparable across the intact and truncated conditions. Paired-sample t-tests show significantly higher rates of boundary perception for intact stimuli indicating that silent pause is an important but not necessary cue to boundary perception and cue redundancy allows for robust perception.

4aSC16. Dichotic digit listening in Mandarin and English by Mandarin-speaking adults. Shu-Yu Liu (School of Speech Lang. Pathol. and Audiol., Chun Shan Medical Univ., Chien-Kuo North Rd., Taichung 402, Taiwan, audio@csmu.edu.tw) and Jia-Shiou Liao (Chun Shan Medical Univ., Taichung 402, Taiwan)

This study examines Mandarin speakers' performance on Mandarin and English dichotic digit recognition tests (DDTs). The 60 right-handed subjects, whose primary language was Taiwan Mandarin, had started English as their second language no later than in seventh grade and continued with English through their freshman year in Taiwan. All the subjects were tested on their ability to pronounce and use English digits before participating in the experiments. Subjects took Mandarin one- and two-pair DDTs, and English one- and two-pair DDTs, in free-recall paradigms. In each DDT, subjects were asked to report the digits orally in any order no matter in what order they had heard the numbers in each ear. The scores were calculated from the number of digits the participants responded correctly to in each ear on one- or two-pair tests; they were then statistically analyzed. The English one- and two-pair recognition tests showed a significant right-ear advantage (REA) but not the Mandarin ones. The scores on the Mandarin one- or two-pair DDTs are higher than those on the English ones, suggesting REA or the influence of the non-native language on the subjects' performance.

4aSC17. Lexical recognition memory across dialects. Cynthia G. Clopper (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210, clopper.1@osu.edu) and Terrin N. Tamati (Indiana Univ., Bloomington, IN 47405)

Implicit recognition memory for spoken words is more accurate when words are repeated by the same or a similar talker than when they are repeated by a different talker. The current study explored implicit recognition memory for words repeated by the same talker, by a different talker from the same dialect, and by a different talker from a different dialect in a word recognition task in noise. Repetitions produced by the same talker facilitated word recognition performance. However, for target words originally pro-

duced by talkers from the Northern dialect of American English, repetitions produced by a talker from the Midland dialect of American English inhibited word recognition performance. No repetition effect was observed for repetitions produced by a different talker from the same dialect or for words repeated by Northern talkers that were originally produced by Midland talkers. These results suggest an asymmetry in how indexical information is stored and activated in lexical processing. The same talker repetition effect was observed for both dialects, but one variety (the Northern dialect of American English) inhibited the repetition effect across dialects and the other variety (the Midland dialect of American English) did not.

4aSC18. Perceptual similarity of unfamiliar regional dialects. Terrin N. Tamati (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405, ttamati @indiana.edu)

Linguistic experience has been shown to influence the perception of regional dialect variation. Recent studies have found that the amount and type of experience with regional dialect variation affect performance in identifying or categorizing regional dialects. Experience also shapes the perceived similarity among regional varieties. To examine the effect of familiarity on the perceived similarity of regional dialects, a paired comparison perceptual similarity rating task was carried out with a group of unfamiliar regional dialects. Native speakers of American English made explicit judgments about the similarity of unfamiliar talkers from the United Kingdom and Ireland based on the regional dialect. Results show that listeners judged the regional dialects of pairs of talkers from the same dialect region as more similar than those of pairs of talkers from different dialect regions. A multidimensional scaling analysis revealed two dimensions of perceptual dialect similarity, both reflecting the geographic location of the cities of origins of the talkers (north versus south and east versus west). Thus, despite being unfamiliar with the regional dialects in the study, listeners were able to use dialect-specific differences in the acoustic signal to make judgments on the perceptual similarity of talkers based on regional dialect.

4aSC19. Identification of the place of articulation of trilingual postvocalic nasals and stops by native speakers of American English, Korean and Japanese. Takeshi Nozawa (College of Economics, Lang. Edu. Cntr., Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan) and Sang Yee Cheon (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Native speakers of American English and Korean produced postvocalic nasals and stops in /CVC/ frames in which the syllable final segment was a consonant with airstream blocking, and native speakers of Japanese produced more nasal /N/ and obstruent /Q/ in /CVNCV/ and /CVQCV/ frames. In Japanese stimuli, the consonant after the nasal or the obstruent was always a stop. Their utterances were recorded and edited to be used as stimuli for the experiment. The release burst of the English stimuli and the second syllable of Japanese stimuli were deleted. Native speakers of these three languages were recruited as listeners. They identified the place of articulation of the syllable-final nasals and stops of the three languages. As predicted, Japanese listeners performed most poorly because there are no phonemic contrasts between postvocalic nasals or stops in Japanese. Korean listeners outperformed the other two groups of listeners in identifying the place of articulation. Postvocalic stops in Korean are not released, so the Korean listeners may not depend on the release burst to identify the place of articulation of a syllable final stop. However, they made more voicing errors than American listeners probably because voiced stops in Korean cannot occur in a postvocalic position.

4aSC20. Why [spa] not [psa]? On the perceptual salience of initial /s/-stop and stop-/s/ sequences. Asimina Syrika (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, Goodnight Hall, 1975 Willow Dr., Madison, WI 53706, syrika@wisc.edu), Jan Edwards, Marios Fourakis, Eun Jong Kong (Univ. of Wisconsin-Madison, Madison, WI 53706), Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455), and Mary E. Beckman (Ohio State Univ., Columbus, OH 43210)

Initial /s/-stop clusters occur frequently in the world's languages, but initial stop-/s/ clusters are relatively infrequent. Furthermore, there appear to be no languages that contain initial stop-/s/ clusters, but not /s/-stop clusters, while the reverse is not true [Morelli, (1999) and (2003)]. This study aims at uncovering a perceptual explanation for these patterns by examining the salience of initial /s/-stop and stop-/s/ clusters in Greek, where both sequences are common. Twenty naïve Greek adult listeners identified syllables beginning with /sp/, /st/, /sk/, /ps/, /ts/, or /ks/, in two vowel contexts, /a/ and /i/, in real words spoken by ten Greek adult native speakers. The syllables were mixed with parts of Greek multitalker babble using SNRs of −6, 0, and +6 dB and presented to listeners for identification. Results showed significantly poorer identification for the /ps/ and /ks/ clusters than the /ts/ and /s/-stop clusters, particularly in the −6 and 0 SNRS. There was also a significant interaction with vowel, such that /sk/ and /ts/ were identified more accurately before /a/, whereas /ks/ was identified more accurately before /i/. The implications of these findings for phonological acquisition and speech perception are considered. [Work supported by NIDCD 02932 and NSF Grant 0729140.]

4aSC21. The perceptual acquisition of Korean fricatives by first language Mandarin listeners. Jeffrey J. Holliday (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, jeffh @ling.ohio-state.edu)

Despite numerous studies, it remains unclear how naive "foreign language" listeners become proficient L2 listeners, particularly in regard to difficult L2 contrasts. In an earlier study in which nine English-speaking L2 learners of Korean of varying proficiency and history of exposure to Korean identified Korean tense /s*/ versus non-tense /s/, listeners showed varying perceptual strategies. The results suggested that learners gradually learn to perceive differences in L2 contrasts by re-weighting useful cues and learning to ignore the "inefficient" cues that are initially relied on when the members of the L2 contrast are assimilated to L1 categories. This paper will report on the results of the same task (identification of CV sequences excised from real words), testing 30 L1 Mandarin speakers who have been in an intensive Korean language program in Seoul for about 2 months. The results of the present study will show whether there are as many inter-listener differences when the level and type of L2 exposure are much more controlled. In addition, because the acoustic cues relevant to the Mandarin sibilant fricative distinctions differ from those used in English the results will show to what extent the choice of perceptual strategy depends on acoustic properties of L1 contrasts.

4aSC22. Cross-linguistic perception of velar and alveolar obstruents: A perceptual and psychoacoustic study. Timothy Arbisi-Kelm (Dept. of Commun. Sci. and Disord., Augustana College, 639 38th St., Rock Island, IL 61201, timothyarbisi-kelm@augustana.edu), Jan Edwards (Univ. of Wisconsin-Madison, Madison, WI 53706), and Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455)

It is well-known that velar stop consonants coarticulate more with the following vowel than stops at other places of articulation. The fine phonetic detail of this coarticulation is highly language-specific. For example, /k/ in Greek is more front before front vowels and more back before back vowels relative to /k/ in English [Arbisi-Kelm *et al.* (2008)]. The purpose of this study was to investigate how these cross-linguistic differences in production influence perception of place of articulation for lingual stops. The stimuli were word-initial consonant-vowel (CV) sequences excised from words produced by 2- to 5-year-old children and adults. The listeners were 20 adult native English speakers (tested in Minneapolis, USA) and Greek speakers (tested in Thessaloniki, Greece) who listened to these sequences combined across ages and languages in a visual analog scaling task [Urberg-Carlson *et al.* (2008)]. Listeners rated how alveolar or velar each sequence was by clicking on a double-headed arrow anchored with language-specific orthographic representations of the target consonants. Results showed that the two groups of adults perceived the sounds differently, as would be expected. We will report on the relationship between listeners' perception and psychoacoustic properties of the stop bursts. [Work supported by NIDCD 02932 and NSF BCS072914 and BCS0729277.]

4aSC23. Individual differences in use of English fricative perceptual cues. Elizabeth Casserly (Dept. of Linguist., Indiana Univ., Memorial Hall 322, Bloomington, IN 47405, casserly@indiana.edu)

This study examines individual variation as a potential explanation for contradictory and inconsistent reports of English speakers' use of acoustic cues in identification of the voiceless sibilants [s] and [ʃ]. While there is widespread agreement that the spectral shape of turbulent noise is key for

identification of these two categories, some studies find that formant transitions to and from the noise also influence identification [Whalen, J. Acoust. Soc. Am. **69**, 275–282 (1981)], while others do not [Harris, Lang. Speech **1**, 1–7 (1958)]. Similarly, the majority of studies investigating the effect of vowel context on fricative perception show that the presence of round vowels biases listeners toward perception of [s] [Kunisaki and Fujisaki, Ann. Bull. RILP **11**, 85–91 (1977)], but others show precisely the opposite effect, where round vowels favor [ʃ] responses [Nittrouer and Studdert-Kennedy, J. Speech Hear. Res. **30**, 319–329 (1987)]. In this study, 30 native English speakers participated in a labeling experiment that fully crossed all three factors—spectral noise shape, formant transitions, and vocalic context—for each subject. Every pattern of cue use found in the literature is also found in one or more of the individuals, which may explain why averaged results vary so widely across reports.

4aSC24. Learning and generalization of novel contrasts across speakers. Kyuwon Moon (kyuwon@stanford.edu) and Meghan Sumner (Dept. of Linguist., Stanford Univ., Stanford, CA 94305-2150)

This paper examines how listeners use a learned contrast when encountering novel speakers. Do speakers reset to their native perceptual biases or apply a learned contrast to new speakers? In two experiments, participants took a minimal pair decision pre-test (MPD), a training session in which a native Korean speaker contrasts stop release (e.g., [bEt] = BET, [bEt'] = BED) without any V/C duration differences, a post-test (identical to pre-test), and a generalization-test [identical to pre-test, but with a speaker of a different L1 (Arabic)]. In Experiment 1, the only difference between the post-test and gen-test was the L1 of the speakers. We found that a learned phonetic contrast generalizes across speakers of different L1s with equally strong effects for post-test and generalization-test independent of order of presentation. In Experiment 2, the post-test was identical to that in Experiment 1, but the learned contrast was paired with vowel durations consistent with native English. Listeners were slower and less accurate in both the generalization-test and post-test when the generalization-test was presented before the post-test. The resulting asymmetry between the two experiments suggests that listeners use learned contrasts, but quickly reset to native patterns when native cues are present.

4aSC25. Perceptual learning of talker-idiosyncratic phonetic cues. Alexandra Jesse (Max-Planck-Institut für Psycholinguistik, Postbus 310, 6500 AH Nijmegen, The Netherlands, alexandra.jesse@mpi.nl) and Rochelle S. Newman (Univ. of Maryland, College Park, MD 20742)

A number of recent studies have explored "perceptual learning," in which listeners use lexical knowledge to learn about a talker's idiosyncratic phoneme pronunciations and adjust their perception of other tokens from that talker accordingly. In a typical perceptual learning study, listeners might hear an item that is ambiguous between "crocotile" and "crocodile" during exposure. Since only crocodile is a word, listeners would learn (following several examples) that this talker has long VOTs, and subsequently at test show a shift in their categorization of a /d/-/t/ VOT continuum by the same talker. The present study explored perceptual learning through cues rather than through lexical knowledge. We used a phonetic contrast (s-th) in which there are both primary (spectral) and secondary (amplitude/duration) cues to phonetic identity. Listeners heard tokens of minimal s-th word pairs in which either the primary or secondary cue was ambiguous, but the alternative cue was unambiguous and thus disambiguated the phonetic identity of the word. We tested whether listeners use the unambiguous cue to learn about the speaker's production of the ambiguous cue (even though doing so was unnecessary for lexical identification) which would then influence later perceptual identification of a series based only on that cue.

4aSC26. Talker-specific accent: Can speech alignment reveal idiolectic influences during the perception of accented speech? Rachel M. Miller, Kauyumari Sanchez, Lawrence D. Rosenblum, James W. Dias, and Neal Dykmans (Dept. of Psych., Univ. of California, 900 Univ. Ave., Riverside, CA 92521, rmiller.ucr.grad@gmail.com)

Listeners use talker-specific (idiolectic) information to help them perceive and remember speech [e.g., Goldinger, J. Exp. Psychol. Learn. **22**, 1166–1183 (1998)]. However, recent research has shown that idiolectic information is not as helpful when listeners hear accented speech [e.g., Sidaras *et al.*, J. Acoust. Soc. Am. **125**, 5 (2009)]. It could be that listeners fail to encode idiolectic information when perceiving accented speech. To examine whether idiolectic is still encoded, experiments tested if subjects would display speech alignment to specific accented models. Speech alignment is the tendency to imitate another talker and can occur when shadowing heard speech [e.g., Goldinger, Psychol. Rev. **105**, 251 (1998)]. Native English subjects were asked to shadow a Chinese- or Spanish-accented model producing English words. Raters then judged whether the shadowed tokens were more similar in pronunciation to those of a shadowed model or of a different models with the same accent. In a second experiment, raters judged whether shadowed tokens were more similar in accent to those of (unshadowed) models with the same or a different accent. Preliminary results reveal that subjects align to the shadowed model, suggesting that idiolect is still encoded. Subjects also show moderate alignment to accent.

4aSC27. Initial acoustic-phonetic processing of competing verbal stimuli examined using dichotic verbal transformations. Peter W. Lenz, James A. Bashford, Jr., and Richard M. Warren (Dept. of Psych., Univ. of Wisconsin-Milwaukee, P.O. Box 413, Garland 224, Milwaukee, WI 53201, plenz@uwm.edu)

Initial studies with dichotic verbal transformations (VTs) of repeating words employed a cross-ear asynchrony of half the word's duration and listeners called out the independent perceptual changes at each ear as they occurred. In contrast to monaural and diotic VTs, the dichotic "immediate response" procedure is extremely difficult and tiring, requiring concurrent monitoring while remembering the word previously heard on each side. The present study employs a much less demanding task—a cued report procedure: the listener calls out what is heard on each side when prompted by a periodic light flash. When the asynchronous stimuli were statements of the same word, the transition rates on each side were the same as when presented monaurally, in contrast with the decreased rates reported with the earlier procedure. With different words on each side, the transition rates were diminished by an amount depending on the extent of their phonetic differences—rates were not influenced by either semantic relations or differences in the neighborhood density of the competitors. It is suggested that dichotic verbal transformations provide access to aspects of the acoustic-phonetic front end of speech analysis that may be obscured by subsequent levels of processing. [Work supported by NIH.]

4aSC28. PRESTO: Perceptually robust English sentence test: Open-set—Design, philosophy, and preliminary findings. Hanyong Park (Speech Res. Lab., Univ. of Indiana, 1101 E. 10th St., Bloomington, IN 40405, hanypark@indiana.edu), Robert Felty (Univ. of Colorado, Boulder, CO), Kelly Lormore (Indiana Univ. Cochlear Implant Program, Indianapolis, IN), and David B. Pisoni (Univ. of Indiana, Bloomington, IN)

Traditional clinical word recognition tests (e.g., SPIN and HINT) do not reflect variability in real world environments. These tests are often constructed with predictable short sentences spoken by a small number of talkers without dialect variation; thus, individual differences are difficult to uncover. To compensate for these shortcomings, PRESTO uses materials taken from the TIMIT database in order to incorporate variation in talkers, dialects, and number of words in a sentence. Familiarity and lexical frequency, syntactic structure, and semantic content are also considered. To provide normative data, normal-hearing young listeners performed an open-set identification task with the PRESTO materials in open field with six-talker babble mixed at different signal-to-noise ratios and also with no additional background noise. The HINT sentences were also used with the same listeners to assess the validity of PRESTO. Results indicate that listeners perform near ceiling for both the PRESTO and HINT materials in ideal listening conditions. However, under degraded listening conditions, the PRESTO shows much greater variability. Preliminary results from a clinical population also indicate that the more challenging PRESTO lists reveal individual differences among patients which are not apparent from scores on the HINT. [NIDCD T32-DC00012.]

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, and Architectural Acoustics: Maximum Entropy and Bayesian Signal Processing I

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., Architecture, 110 8th St., Troy, NY 12180

Zoi-Heleni Michalopoulou, Cochair
New Jersey Inst. of Technology, Dept. of Mathematics, Newark, NJ 07102-1982

Chair's Introduction—7:35

Invited Papers

7:40

4aSP1. Bayesian approach to model-based signal processing: An overview. James V Candy (Lawrence Livermore Natl. Lab., P.O. Box 808, L-151, Livermore, CA 94551)

Although available for a long time with the advent of high-speed/high-throughput computing, the development of Bayesian processing techniques has evolved recently in acoustical signal processing. Bayesian signal processing is concerned with the estimation of the underlying probability distribution of a random signal in order to perform statistical inferences such as the conditional mean estimation. Knowledge of this distribution provides all of the essential information available required for problem solution. The usual limitations of nonlinear approximations and non-gaussian processes prevalent in classical algorithms (e.g., Kalman filters) are no longer a restriction to perform Bayesian inference. This approach enables the next generation of processors called particle filters that are sequential Monte Carlo methods providing an estimate of the underlying discrete probability distribution. In this overview, Bayesian signal processing is presented from a probabilistic perspective starting with Bayes rule and evolving to the development of a bootstrap particle filter perhaps one of the most common and simplest constructs available. The relationship of Bayesian processing to the concept of maximum entropy is discussed. Maximum entropy and its applicability in Bayesian processing is also mentioned briefly.

8:00

4aSP2. Defining uncertainty with maximum entropy method. David P. Knobles, Jason Sagers, and Robert Koch (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

The maximum entropy (ME) method was strongly defended and advocated by E. T. Jaynes as a means to define uncertainty. Here, the ME method is applied to the estimation of ocean waveguide parameter probability distributions from measured acoustic data. An ME analysis produces a canonical distribution, well known from equilibrium statistical mechanics, which is the distribution that maximizes the entropy subject to constraints that reflect selected features of the measured data and a model. The ME method gives the most conservative distribution based only on the measured data and observed features. A Bayesian approach also has the goal of defining uncertainty, but starts from the specification of the likelihood function and the model priors. Data noise is naturally handled in the specification of the likelihood function. The discussion introduces simple examples, showing basic relationships between the constraints in ME and the maximum likelihood estimation. In special cases the form of the likelihood function used in Bayesian conditionalization can be derived from the ME approach. The form of the cost function is an important consideration in comparing ME and Bayesian methods of inferences. In general ME and Bayesian inferences lead to different results. [Work supported by ONR Code 321 OA.]

8:20

4aSP3. Automatic signal detection in noise using entropy. Christine Erbe (JASCO Appl. Sci., Brisbane Technol. Park, 1 Clunies Ross Ct, Eight Mile Plains, QLD 4113, Australia, christine@jasco.com)

Automatic detection of signals in noise is a common problem in many areas of acoustics. In the field of passive acoustic monitoring of marine mammals, the signals to be detected are vocalizations. The noise originates from natural (wind, waves, and rain) and man-made sources (e.g., shipping, construction, and seismic surveys). Signal characteristics vary broadly: frequency ranges from a few Hz to 200 kHz, duration from milliseconds to seconds to hours. Noise characteristics vary by similar orders of magnitude. While specific automatic detectors have been designed to successfully find specific calls in specific environments, the challenge is to find a large variety of calls in a large variety of noise. An exploitable difference between calls and noise is that most noise is a result of stochastic processes (wind, waves, rain, cavitating propellers + seismics generate gas bubbles underwater of varying size + resonance frequency), while many animal signals are a result of deterministic processes (vibrating strings & cavities of predetermined/fixed size). Shannon entropy was computed for power spectrum density functions of underwater recordings. Noise yielded high signal low entropy. Results are presented from passive acoustic surveys of marine mammals. The benefits and limitations of entropy applied to automatic signal detection are discussed.

8:40

4aSP4. The likelihood ratio and Bayesian signal processing. R. Lee Culver and Colin W. Jemmott (Appl. Res. Lab and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

Over the past several years, we have been developing an architecture for classifying source depth associated with passive sonar signals. The classifiers utilize the statistics of a signal parameter which are estimated using knowledge of the environment and an acoustic propagation program. We have applied the likelihood ratio (LR) test to classify source depth using signal statistics from the SWellEx-96 and 1996 Strait of Gibraltar sea tests. More recently Bissinger developed a Hellinger distance classifier, and Jemmott is developing a histogram (discrete Bayesian) filter for this purpose. In this talk, we examine the relationship between the LR test and a processor that makes use of Bayes rule. We consider some of the fundamentals. It is useful to understand the underlying assumptions of the LR, the likelihood function, and how they are related to a Bayesian processor which makes use of prior information and computes a posterior probability distribution function. Under what conditions do the two processors produce the same answer? When would the Bayesian processor be a better choice? We compare the processors and apply them to the SWellEx-96 data. [Work supported by the Office of Naval Research Undersea Signal Processing.]

9:00

4aSP5. Bayesian bounds on passive sonar accuracy from binary performance metrics. John R. Buck (Dept. ECE, Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., N. Dartmouth, MA 02747, johnbuck@ieee.org)

A passive sonar algorithm measures the pressure field at a sensor array then estimates the sound source location from these observations and an acoustic propagation model. Passive sonar performance is usually characterized by the mean squared error (MSE) between the estimated and true source locations. Consequently, passive sonar performance bounds typically provide lower bounds on the achievable MSE for a given array and environment. Such MSE bounds can be misleading in environments with strong sidelobes, as the optimal estimator may choose an unlikely location to balance the error among several highly likely locations. An alternative algorithm would be to partition the search space, then use the array observations to choose which block contains the source, but not its exact location in the block. The resulting binary performance metric is now the probability of choosing the incorrect block or error probability (P_e). Information theory allows us to formulate Bayesian bounds on the minimum achievable P_e for a given array, propagation environment, and search space partition. These bounds quantify the trade-off among SNR, P_e, and location estimation accuracy for passive sonar. [Work supported by ONR Code 321US.]

9:20

4aSP6. Using Bayesian inference for acoustic array design. Paul M. Goggans and Chung-Yong Chan (Dept. of Elec. Eng., Univ. of Mississippi, Anderson Hall, Rm. 302B, University, MS 38677)

Because inference and design are both generalized inverse problems, the tool and methods developed for Bayesian parameter estimation and model comparison can be adapted and used for the solution of design problems. As an example, this paper presents the use of the Bayesian inference framework for the automated design of linear transducer arrays [P. M. Goggans and C.-Y. Chan, "Antenna array design as inference," AIP Conf. Proc. **1073**, 294–300 (2008)]. Commonly, automated array design is cast as an optimization problem and solved using numerical optimization techniques. Here, the design of linear arrays is cast as an inference problem and solved using numerical Bayesian inference techniques. Compared to optimization-based methods, the inference-based method presented here has the advantage of being able to automatically determine the number of array elements required to satisfy design requirements and specifications. In addition, array design cast as inference can incorporate, as prior information, design requirements such as a minimum spacing between two adjacent elements, a maximum aperture width, and a necessary operating frequency bandwidth. Sample results are presented to demonstrate the application of the Bayesian inference framework in the automated design of linear arrays.

9:40

4aSP7. Recursive Bayesian state estimation for passive sonar localization. Colin W. Jemmott and R. Lee Culver (Penn State Appl. Res. Lab and Grad. Program in Acoust., P.O. Box 30, State College, PA 16804, cwj112@psu.edu)

A model-based recursive Bayesian signal processing framework is shown to localize a moving source emitting a low-frequency tonal signal in a shallow water environment. Source motion maps spatial variation in transmission loss into amplitude modulation of the signal received on a passive horizontal array. Acoustic propagation modeling predicts this variability, which is used to estimate source range, depth, range rate, and acoustic level. Uncertainty in transmission loss resulting from uncertainty in environmental parameters is predicted using Monte Carlo modal propagation modeling. Monte Carlo marginalization over environmental uncertainty provides robustness against data-model mismatch. The maximum entropy method is used to construct a probability density function (pdf) of transmission loss at each range depth location based on the Monte Carlo results. The resulting pdfs belong to the exponential family and result in an implementable recursive Bayesian processor. The physics of acoustic modeling determine the form of the processor through the transmission loss pdfs and are an intimate part of the localization technique. This processor is distinct from Bayesian matched field processing in that it neither relies on a vertical array nor computes modal amplitudes from received data. Results using SWellEx-96 will be shown. [Work supported by ONR Undersea Signal Processing.]

10:00—10:15 Break

10:15

4aSP8. Bayesian geoacoustic inversion. Stan E. Dosso and Jan Dettmer (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper describes a general Bayesian approach to estimating seabed geoacoustic parameters from ocean acoustic data, which is also applicable to other inverse problems. Within a Bayesian formulation, the complete solution is given by the posterior probability density (PPD), which includes both data and prior information. Properties of the PPD, such as optimal parameter estimates, variances /covariances, correlations, and marginal probability distributions, are computed numerically for nonlinear problems using Markov-chain Monte Carlo methods. However, in many practical cases, both an appropriate model parametrization and the data error distribution are unknown and must be estimated as part of the inversion. These problems are linked, since the resolving power of the data is affected by the data uncertainties. Model selection is carried out by evaluating Bayesian evidence (parametrization likelihood given the data), or a point estimate thereof such as the Bayesian information criterion, which provides the simplest parametrization consistent with the data. The error covariance matrix (including off-diagonal terms, as needed) is estimated from residual analysis under the assumption of a simple, physically reasonable distribution form, such as a Gaussian or Laplace distribution. The validity of the above assumptions and estimates is examined *a posteriori* using both qualitative and quantitative statistical tests.

10:35

4aSP9. Particle filtering for sequential multipath arrival time and amplitude estimation. Rashi Jain and Zoi-Heleni Michalopoulou (Dept. of Math. Sci., New Jersey Inst. of Technol., Newark, NJ 07102)

Accurately estimating arrival times from acoustic time series in the ocean is essential for successful source and array element localization and estimation of the geometry of the sound propagation environment and environmental parameters such as sound speed in the water column and sediments. We have developed a sequential Monte Carlo method that characterizes multipath arrivals as moving targets, tracking them at spatially separated receiving phones. We focus on switching models that are suitable for unknown and varying numbers of arrivals at different phones. We also present approaches that efficiently and effectively extract amplitude information from received time series; such information can be then employed for sediment characterization. Our methods are applied to Haro Strait Primer and Shallow Water 06 data; their performance is evaluated through comparisons to conventional approaches. [Work supported by ONR.]

Contributed Papers

10:55

4aSP10. Sequential Bayesian strategies in geoacoustic inverse problems. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada), and Charles W. Holland (The Penn State Univ., State College, PA)

This paper considers sequential Bayesian strategies for geoacoustic inverse problems which are difficult to solve simultaneously due to computational constraints. Bayesian inference provides a powerful approach to learning problems such as this since sequential inversions of multiple data sets [with the posterior probability density (PPD) of one inversion applied as prior information in the subsequent inversion] are equivalent to simultaneous inversion of all data. However, passing PPDs forward as priors has its own challenges when the PPD is sampled numerically for nonlinear inverse problems, particularly when the model parameter space is of high dimensionality and the data information content is high. In such cases, approximations are required to efficiently carry PPD information forward to subsequent inversions. The approach developed here represents numerically sampled PPDs in terms of discretized marginal probability distributions for principal components of the parameters, which minimizes the loss of information in representing inter-parameter correlations. The sequential Bayesian approach is applied to seabed reflectivity inversion with multiple data sets representing travel-time data and frequency-domain reflection coefficient data for a series of increasing penetration depths. Data information content is quantified by accounting for potential error biases as well as data error covariances. [Work supported by the Office of Naval Research.]

11:10

4aSP11. Three-dimensional source tracking in an uncertain environment via Bayesian marginalization. Dag Tollefsen (Norwegian Defence Res. Est. (FFI), Box 115, 3191 Horten, Norway) and Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper develops a non-linear Bayesian marginalization approach for three-dimensional source tracking in shallow water with uncertain environmental properties, with application to horizontal line array (HLA) data. The algorithm integrates the posterior probability density via a combination of Metropolis–Hastings sampling over environmental and bearing model parameters and Gibbs sampling over source range and depth, with a priori track constraints on source velocity. Two-dimensional marginal distributions for source range/depth and range/bearing are derived. The Viterbi algorithm is applied to obtain the most probable track, with uncertainties estimated from the marginal distributions. The algorithm is applied to simulated data in continental shelf environment and to towed-source and ship-noise data recorded on a HLA deployed on the seafloor in an experiment conducted in the Barents Sea.

11:25

4aSP12. Computation of normalizing constants in geoacoustic Bayesian inference. Jan Dettmer and Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper considers approaches to computing normalizing constants (Z) in Bayesian inference problems. Bayes' theorem combines the likelihood function, model prior, and Z to form the posterior probability density (PPD). Z (also known as evidence) is difficult to compute for general problems and a common approach is to avoid its computation entirely by calculating an unnormalized estimate of the PPD which is sufficient for moment estimates. However, estimating the normalized PPD, including Z, allows for moment estimates as well as quantifying the likelihood of the model parametrization. This is commonly referred to as model selection and poses a natural way to quantifying the most appropriate model parametrization for a given data set (Bayesian razor). Several approaches for computing Z have been developed in the statistics community, some of which are applied here to the geoacoustic inference problem. Annealed importance sampling follows an annealing approach and computes weighted averages along cooling trajectories. Nested sampling uses a likelihood constraint to move from the prior mass to the posterior. Both methods also give parameter estimates which are compared to Metropolis–Hastings results. [Work supported by the Office of Naval Research.]

Session 4aUW

Underwater Acoustics: Propagation and Scattering in Heteorogeneous Waveguides

Jon Collis, Chair
Colorado School of Mines, Dept. of Mathematical and Computer Science, 1500 Illinois St., Golden, CO 80401

Contributed Papers

8:00

4aUW1. Improving the parabolic equation solution for problems involving poro-elastic media. Adam M. Metzler, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180), Michael D. Collins, Ralph N. Baer (Naval Res. Lab., Washington, DC 20375), and Jon M. Collis (Colorado School of Mines, Golden, CO 80401)

Parabolic equation solutions for elastic media have undergone several modifications recently that increase their capabilities and accuracy. These include formulation in different dependent variables, approaches for handling range dependence such as coordinate rotations and single scattering, and treatment of media anisotropy. These advances are being extended to problems with heterogeneous and range-dependent poro-elastic media, which provide useful models of some shallow-water sediments. Other parabolic equation solutions for poro-elastic media [Collins *et al.*, J. Acoust. Soc. Am. **98**, 1645–1656 (1995)] are prior to recent progress. Vertical dependence is treated by applying heterogeneous depth operators from the equation of motion. Horizontal dependence is treated by incorporating single-scattering approaches. [Work supported by the ONR.]

8:15

4aUW2. Seismo-acoustic propagation near low-shear speed poroelastic ocean sediments using a hybrid parabolic equation solution. Jon M. Collis (Colorado School of Mines, Golden, CO 80401)

Accurate and efficient parabolic equation solutions exist for complex propagation environments featuring elastic and porous elastic sediment types. An area of concern has been low-shear wave speed sediments that become singular as their shear modulus tends toward zero. A historic approach for treating sediments of this type has been to assume that it is a fluid, and effects due to elasticity are negligible. This approach is limited in accuracy unless shear is accounted for. In this presentation, the ocean bottom sediment interface layer is treated as a porous elastic layer in which poroelastic momentum equations are solved and combined with an existing elastic parabolic equation implementation. Appropriate boundary conditions are enforced at the fluid-poroelastic and poroelastic-elastic interfaces. The new solution is tested on problems with a low-shear ocean bottom interface layer.

8:30

4aUW3. Improving the parabolic equation solution for problems involving sloping fluid-solid interfaces. Michael D. Collins (Naval Res. Lab., Washington, DC 20375, collins@noddy.nrl.navy.mil) and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, New York 12180)

Several approaches are being investigated for improving the elastic parabolic equation for problems involving sloping fluid-solid interfaces. Approaches based on single scattering and energy conservation provide accurate solutions for problems involving sloping fluid-fluid and solid-solid interfaces, but the mixed-media problem has proven to be more challenging. The energy-conservation approach has been applied previously by deriving a linear equivalent to the nonlinear expression for energy flux. One of the approaches that are currently being investigated is based on going back to the nonlinear expression. Although it would not be practical to solve the full nonlinear scattering problem, promising results have been obtained for the fluid-fluid case with this approach by correcting the amplitude at only one grid point near the interface. With this approach, the nonlinear problem reduces to the evaluation of a quadratic function. Another approach that is being investigated is based on an alternative formulation that involves the vertical displacement and a quantity that is proportional to the normal stress on a horizontal interface. In these variables, the interface conditions across a horizontal interface are first order, and this may facilitate the extension of the single-scattering solution to the mixed-media problem. [Work supported by the Office of Naval Research.]

8:45

4aUW4. Dependence of the structure of the shallow convergence zone on deep ocean bathymetry. Stephen D. Lynch (slynch@mpl.ucsd.edu), Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. Oc., San Diego, CA), Kevin Heaney (OASIS, Lexington, VA), Arthur B. Baggeroer (MIT, Cambridge, MA), Peter Worcester (Scripps Inst. Oc., La Jolla, CA), and James Mercer (APL/UW, Seattle, WA)

During an experiment in the northern Philippine Sea in 2009, low-frequency tones were transmitted from a shallow (15- and 60-m) source deployed from R/V Melville keeping station to a shallow (250-m) horizontal receiver array towed by R/V Kilo Moana approximately one convergence zone (CZ) away. Recordings were made during events in which the receiver ship maintained constant range in the convergence zone and during events in which the receiver ship transited radially through the CZ. The shallow CZ exhibits strong dependence on the bathymetry mid-way between the source and receiver array. In fact, the variability of the structure of the first CZ in this environment is significantly more strongly affected by the heterogeneous character of the bottom than water column fluctuations. Numerical modeling with a parabolic equation code is used to support the conclusions from the data analysis.

9:00

4aUW5. Range dependence in the level set method for underwater acoustics. Sheri L. Martinelli (Div. of Appl. Mathematics, Brown Univ., 182 George St., Providence, RI 02912)

The level set method due to Osher and Sethian [J. Comput. Phys. **79**, 12–49 (1988)] provides a way to obtain fixed grid solutions to the high-frequency wave equation. Instead of tracing rays from the source, the level set method embeds the wavefront implicitly in the phase space and propagates it according to the velocity field determined by the local ray direction, thus avoiding the complications involved in the spatial reconstruction of wavefronts from diverging rays. A level set method has been developed and implemented as a fixed-grid alternative to ray tracing to solve for the acoustic phase. One of the issues that arises with the increased dimensionality of posing the propagation problem in the level set framework is that the presence of reflecting boundaries produces a discontinuity in the phase space corresponding to a sudden change in propagation direction. When a reflecting boundary is range-dependent, further complications arise. To improve algorithm performance, specialized methods are applied to the level set equations that combine upwinding with higher-order spatial interpolation that avoid the generation of spurious oscillations that occur with most traditional finite difference methods. [Work supported by ONR 333 and the SMART Program.]

9:15

4aUW6. Two dimensional finite element propagation and reverberation modeling in shallow water waveguides. Marcia J. Isakson (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

Finite element propagation models do not rely on approximations of the scattering at the interfaces and therefore provide excellent benchmark solutions to rough interface waveguide propagation and reverberation studies. In this study, two dimensional finite element solutions for reverberation and propagation are calculated for waveguides that have rough interfaces at the air/water boundary and the water/sediment boundary. The effects of upward and downward refracting sound speed profiles are also considered. [Work sponsored by Office of Naval Research, Ocean Acoustics.]

9:30

4aUW7. An algorithm to predict surface loss. Cathy Ann Clark (NUWCDIVNPT, B1320, R457, 1176 Howell St., Newport, RI 02841)

A semi-empirical surface loss algorithm is presented which is comprised of a rough scattering component derived from theory and a term which represents low-frequency, low angle loss from other mechanisms such as absorption and bubbles, based on a fit to measured data. A prediction of surface duct propagation using the semi-empirical algorithm and a current Navy standard propagation model is compared to measured data.

9:45

4aUW8. Boundary flattening transforms for acoustic propagation under rough and periodic sea surfaces. Roger Oba (Acoust. Div. Naval Res. Lab., Washington, DC 20375, roger.oba@nrl.navy.mil)

A conformal transform is presented that maps an acoustic domain with a one-dimensional, rough sea surface onto a domain with a flat top. The nonperturbative transform presented here broadly generalizes that of Dozier [J. Acoust. Soc. Am. **75**, 1415–1432] to include many wavelengths of the surface variation. A two-dimensional, flat-top domain permits the direct application of a parabolic equation model acoustic propagation model for the Helmholtz equation using a modified sound speed. Once the field is computed, the inverse transform permits the acoustic field interpolation in terms of the original coordinates. The mapping is derived from techniques in the classical theory of flow around an airfoil. Forward scatter test cases with periodic and rough sea surfaces provide verification of the method using a parabolic equation model. The periodic surface case demonstrates scattering from steep grazing angles to shallower ones. An extension to scattering to irregular cylinders is outlined following the scheme of DiPerna and Stanton [J. Acoust. Soc. Am. **96**, 3064–3076, (1995)]. [This research is sponsored by the Office of Naval Research.]

10:00—10:15 Break

10:15

4aUW9. Three-dimensional scattering from pressure-release rough surfaces. Sumedh M. Joshi and Marcia J. Isakson (Appl. Res. Labs., 10000 Burnet Rd., Austin, TX 78758, sumedhj@mail.utexas.edu)

In order to compare a variety of three-dimensional (3-D) rough surface scattering theories, the scattering of a spherical wave incident on a pressure-release rough surface is modeled. Random surface realizations are computed from a spatial roughness power spectrum measured as part of the EVA sea test conducted in 2006. Scattering from these surfaces is computed using boundary and finite element methods. A singularity removal technique is applied to solve the Helmholtz–Kirchhoff boundary integral equation in 3-D. This integral solution is compared with 3-D finite elements and the 3-D Kirchhoff approximation, to determine the range of validity of the models.

10:30

4aUW10. Boundary roughness and the effect of internal waves on signal coherence for shallow water transmission. Harry, A. DeFerrari (Div. of Appl. Marine Phys., RSMAS - Univ. of Miami, 4600 Rickenbacker Cswy, Miami, FL 33149)

Broadband acoustic propagation experiments at three shallow sites allow for comparison of coherency of individual surface-reflected bottom-reflected modes of propagation. There appears to be a dependence of the correlation parameters of times and length on frequency and mode number that cannot be attributed to internal waves alone and likely depends on bottom and surface roughness. At low frequencies, f$<$100 Hz, during periods of quiescence internal waves, all modes of propagation have equally long coherence parameters. The coherency decrease equally for all modes as internal wave activity increases. For higher frequencies, 200 Hz $<f<$1000 Hz, the coherence parameters depend on mode number, with the lower order modes always more coherent than successive higher order modes. At still higher frequencies, $f>$1000 Hz, identifiable modes are not always observed; instead there is a continuum of arriving pulse energy with very low coherency even with minimal internal waves. Apparently, the randomizing effect of internal waves depends on bottom and surface roughness and frequency. At low frequencies, the boundaries appear flat and internal waves have a minimal effect. At the highest frequencies, phase coherence is already degraded by boundary roughness so that the slightest of internal wave activity completely randomizes the signals.

10:45

4aUW11. Horizontal Lloyd mirrors arising from propagation through straight and curved nonlinear internal wave fronts. Kara G. McMahon, Laurel K. Reilly-Raska (Dept. Math. Scis., Rensselaer Poly. Inst., Troy, NY 12180, mcmahk3@rpi.edu), James F. Lynch, Timothy F. Duda (Woods Hole Ocean. Inst., Woods Hole, MA 02543), and William, L. Siegmann (Rensselaer Poly. Inst., Troy, NY 12180)

Experimental observations and theoretical studies show that nonlinear internal waves (NIWs) occur widely in shallow water and cause acoustic propagation effects including mode coupling and ducting. Horizontal ducting results when acoustic modes interact with NIW fronts that comprise waveguide boundaries. For small grazing angles between a mode trajectory and a front, an interference pattern may arise that is hypothesized [Lynch *et al.*, J. Ocean Eng. **31**, 33–48 (2006)] to be a horizontal Lloyd mirror. We examine acoustic formulations for this feature and benchmark calculations for the acoustic intensity with those from the adiabatic mode parabolic equation. Results using different waveguide features are compared, including continuous-gradient and jump sound-speed profiles of varying strengths. We focus on differences in the location of the source relative to the NIW as well as the frontal curvature. The curvature influences both incidence angles and reflection characteristics. For sources oriented inside the front, as curvature increases the areas with interference patterns shrink, while sources beyond the front cause patterns to expand. [Work supported by ONR.]

11:00

4aUW12. Blind deconvolution of remote-source signals from ocean acoustic array recordings. Shima H. Abadi (shimah@umich.edu), David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, ML 48109), and Daniel Rouseff (Univ. of Washington, Seattle, WA 98105)

Reconstructing the signal originally broadcast from a remote source in an unknown multipath environment is a task commonly known as blind deconvolution. At frequencies of several kilohertz and above, multipath shallow-ocean sound propagation may be adequately described by ray acoustics. This presentation describes results from the application of ray-based artificial time reversal (ATR) to underwater sound propagation measurements. The receiving array was vertical and it recorded signals with center frequencies and bandwidths of a few kHz at source-receiver ranges up to 3 km in a water depth of approximately 60 m. Ray-based ATR uses a simple-beam-former-determined ray-arrival direction to construct a frequency-dependent phase correction at the receiving array that allows the Green's function of the sound channel and the original source waveform to be separately estimated. Here, the correlation coefficient between the original signal and the ATR-reconstructed signal is presented as a function of range and signal-to-noise ratio. In addition, the effect of reducing the number elements of the receiving array and the use of a coherent combination of reconstructed results for various ray arrival directions on cross correlation coefficient will be discussed. [Work supported by ONR.]

11:15

4aUW13. Experimental studies of underwater acoustic communication in Trondheim fjord. Guosong Zhang (Dept. of Electr. and Telecomm., Norwegian Univ. of Sci. and Tech., O.S. Bragstads plass 2B, NO-7491, Norway, guosong.zhang@iet.ntnu.no), Hefeng Dong, and Jens M. Hovem (Norwegian Univ. of Sci. and Tech., NO-7491, Norway)

Direct-sequence spread-spectrum signal was used for communication tests over underwater channels in Trondheim fjord. Differential binary phase shift keying was utilized between two adjacent symbols. To the receiver, a method uses a time updated channel impulse response estimation to recover differential phase modulated information, and it takes the estimation from the previous symbol as the match filter. The effectiveness of this method is ensured by the coherence between two consecutive symbols over time varying channels. This method is insensitive to multipath patterns, and it does not require time synchronization as precise as the convention de-spread method does. In our experiments, good performance was achieved, even in low SNR tests. The performance loss at high SNR in the experiments was caused by long time delay spread. Late-arriving paths from the previous symbol were buried in the current symbol during time-windowing process, and the late-arriving paths might decrease the magnitude of the differential phase information. In this situation, it is prone to cause errors.

11:30

4aUW14. Frequency dispersion of parametric array signal in shallow water. Igor Esipov (N. Andreyev Acoust. Inst., 4, Shvernik str., 117036 Moscow, Russia, igor.esipov@mail.ru), Sergey Tarasov, Vasily Voronin (Inst. of Technol. in Taganrog, Russia), and Oleg Popov (N. Andreyev Acoust. Inst.)

Results of experimental test of parametric array application for marine shallow water waveguide excitation by sweep frequency modulated signal are discussed. Parametrical sound signal is forming in shallow water environment, which is stimulated by intensity modulated high frequency power acoustical pump. As a result the end-fire parametric array is forming there, which excites sharp directional signal radiation at the modulation frequency. Such a low-frequency signal, generated in the virtual end-fire array by parametrical means, will propagate in shallow water waveguide independently from the pump radiation. Shallow water signal propagation obeys to waveguide dispersion. Sweep modulated signal compression is experimentally shown for single mode signal propagation in shallow water. Acoustical signal of 2-ms duration is generated in frequency band of 7–15 kHz by parametric array. This signal is transmitted in single lobe of 2 deg width along the path of 5.6 km long in water layer from 2.5- to 3-m depth. The directivity of the signal transmitted was constant in the whole frequency range. It was shown the single mode excitation of the shallow water waveguide takes place under this circumstance. [Work supported by ISTC, Project No. 3770.]

Meeting of Standards Committee Plenary Group

to be held jointly with the meetings of the
ANSI-Accredited U.S. Technical Advisory Groups (TAGs) for:
ISO/TC 43, Acoustics,
ISO/TC 43/SC 1, Noise,
ISO/TC 108, Mechanical vibration, shock and condition monitoring,
ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures,
ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments,
ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock,
ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,
ISO/TC 108/SC 6, Vibration and shock generating systems,
and
IEC/TC 29, Electroacoustics

P. D. Schomer, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics and ISO/TC 43/SC 1 Noise
Schomer and Associates, 2117 Robert Dr., Champaign, IL 61821

D. J. Evans, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108 Mechanical vibration shock and condition monitoring, and ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring devices
National Institute of Standards and Technology (NIST), 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899

W. C. Foiles, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
BP America, 501 Westlake Park Blvd., Houston, TX 77079

R. Taddeo, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
NAVSEA, 1333 Isaac Hull Ave., SE, Washington Navy Yard, Washington, DC 20376

D. D. Reynolds, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock
3939 Briar Crest Ct., Las Vegas, NV 89120

D. J. Vendittis, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines
701 NE Harbour Terrace, Boca Raton, FL 33431

R. Taddeo, Vice Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines
NAVSEA, 1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington, DC 20376

C. Peterson, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 6 Vibration and shock generating systems
200 Dixie Ave., Kalamazoo, MI 49001

V. Nedzelnitsky
U.S. Technical Advisor (TA) for IEC/TC 29 Electroacoustics
National Institite of Standards and Technology (NIST), 100 Bureau Dr., Gaithersburg, MD 20899–8221

The reports of the Chairs of these TAGs will not be presented at any other S Committee meetings.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S2, S3, and S12 which are scheduled to take place in the following sequence:

ASC S12, Noise	**22 April 2010**	**9:15 a.m. to 10:30 a.m.**
ASC S12, Noise	**22 April 2010**	**9:15 a.m. to 10:30 a.m.**
ASC S2, Mechanical Vibration and Shock	**22 April 2010**	**11:00 a.m. to 12:00 noon**
ASC S1, Acoustics	**22 April 2010**	**1:45 p.m. to 2:25 p.m.**
ASC S3, Bioacoustics	**22 April 2010**	**3:00 p.m. to 4:15 p.m.**
ASC S3/SC1, Animal Bioacoustics	**22 April 2010**	**4:30 p.m. to 5:30 p.m.**

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees and U.S. TAGs.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. Parallel Committee
ISO		
P. D. Schomer, Chair	**ISO/TC 43** Acoustics	ASC S1 and S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	ASC S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical vibration, shock and condition monitoring	ASC S2
W. C. Foiles, Co-Chair R. Taddeo, Co-Chair	**ISO/TC 108/SC2** Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures	ASC S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and calibration of vibration and shock measuring instruments	ASC S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human exposure to mechanical vibration and shock	ASC S2/S3
D. J. Vendittis, Chair R. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition monitoring and diagnostics of machines	ASC S2
C. Peterson, Chair	**ISO/TC 108/SC6** Vibration and shock generating systems	ASC S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	ASC S1 and S3

Meeting of Accredited Standards Committee (ASC) S12 Noise

W. J. Murphy, Chair, ASC S12
NIOSH, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, OH 45226

R. D. Hellweg, Vice Chair, ASC S12
Hellweg Acoustics, 13 Pine Tree Road, Wellesly, MA 02482

Accredited Standards Committee S12 on Noise. Working group chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 43/SC 1 Noise, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 8:00 a.m. on Thursday, 22 April 2010.

Scope of S12: Standards, specifications and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

A. T. Herfat, Chair, ASC S2
Emerson Climate Technologies, Inc., 1675 W. Campbell Road, P.O. Box 669, Sidney, OH 45365-0669

C. F. Gaumond, Vice Chair, ASC S2
Naval Research Laboratory, Code 7142, 4555 Overlook Ave., SW, Washington, DC 20375-5320

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 108, Mechanical Noise, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 8:00 a.m. on Thursday, 22 April 2010.

Scope of S2: Standards, specifications, methods of measurements and test, and terminology in the field of mechanical vibration and shock, condition and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

Session 4pAA

Architectural Acoustics and Engineering Acoustics: Rooms for Reproduced Sound II

K. Anthony Hoover, Chair
McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362

Invited Papers

1:00

4pAA1. Measuring and defining the sound quality of audio systems. Peter Mapp (Peter Mapp Assoc., Colchester, United Kingdom)

"High quality" is a term often employed as a requirement for the performance of an audio or sound system, yet what does it mean and can it be quantified? In particular, bass performance is an important aspect but again is extremely hard to define. Frequency response assessment alone is not enough to characterize sound quality as this does not provide information with regard to the temporal or transient characteristics of the room or system. The paper will look at a number of techniques that potentially can be used to help assess sound system performance including a new measurement method intended to quantify the transient and temporal characteristic of the bass sound in a room, whether this is via a sound system or an assessment of the room itself.

1:20

4pAA2. But how does it look? Deb Britton (K2 Audio, LLC, 4900 Pearl East Circle, Ste. 201E, Boulder, CO 80301, deb@k2audio.com)

As acousticians and audio professionals, we are constantly striving for the "optimum sound solution." Quite often our optimum solution is at odds with the room's architecture or the desired aesthetic. This paper will present some examples of spaces where high-quality sound, even coverage, speech intelligibility and visual aesthetics were all achieved.

1:40

4pAA3. Sweetness spotting. Sam Ortallono (zSam Ortallono MediaTech Inst., 3324 Walnut Bend Ln., Houston, TX 77042)

This paper investigates the traditional listening position for the mixing of stereo music in a near field-monitoring situation. Common sense and current practice reinforce the ideal listening post to be at the apex of an equilateral triangle formed by the two speakers and the observer's head. This experiment will put that notion to the test. Using white noise pink noise, sine waves, and program material we will compare listening positions to the left and right as well as closer and farther. Readings from several mixing studios at multiple facilities will provide the numbers. Results will be correlated with room size, shape, and room treatment. Will the data uphold the longstanding wisdom on phase coherency and stereo image or will the results surprise us all?

2:00

4pAA4. Critical listening research on reproduced sound in diffuse environments. William L. Martens (Faculty of Architecture, Design and Planning, Univ. of Sydney, New South Wales 2006, Australia, bill@arch.usyd.edu.au) and Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

In critical listening rooms designed to monitor precisely small signal processing changes, what can be heard is influenced by the room's modal response and the degree to which reflections are controlled by absorptive, reflective, and diffusive boundary surfaces. In exploring the problem of room acoustics affecting the results of such important listening activities, it is informative to explore the boundary conditions. For example, the acoustics industry utilizes absorption and reverberation chambers to explore related auditory phenomena. Recently, a diffusion chamber was created with broadband diffusion on walls and ceiling. Modal frequencies were also controlled with dedicated plate resonators effective down to 50 Hz. Impulse response measurements in this diffuse chamber revealed a dense reflection pattern 30 dB or more below the direct sound. The initial perception in this room was precise sonic images, a comfortable sense of ambiance and an apparent transparency to the subtleties of reproduced imagery that support decisions of sound engineers. Related experiments being carried out in other controlled listening rooms will be reviewed to place the current work in context.

2:20

4pAA5. How reverberation degrades and aids source perception in auditory scenes. Barbara G. Shinn-Cunningham (Boston Univ. Hearing Res. Ctr., 677 Beacon St., Boston, MA 02215)

Reverberant energy (or reverberation) influences perception of source content (in addition to influencing perception of the environment and of source location). Moreover, reverberation can affect perception in different ways depending on whether a source is played in quiet or in a scene containing competing sounds. Reverberation smears spectro-temporal structure, which can interfere with estimation of duration, pitch contour, and other basic attributes. In the extreme, such spectro-temporal smearing (self-masking) can interfere with extracting the meaning of signals like speech or music. However, if the direct sound reaching a listener is nearly inaudible, re-

verberation adds energy, improving source detectability. While these in-quiet effects contribute to perception of sources in a complex acoustic scene, additional factors also come into play. For instance, if two otherwise similar sources have different levels of reverberation, the reverberation can make them more perceptually distinct, making it easier to focus attention on one and understand it. On the other hand, the same spectro-temporal smearing that can hurt perception of a source in quiet can also disrupt source segregation, making it difficult to selectively analyze a desired source. Examples of these different effects of reverberation on perception, taken from recent experiments, will be presented and discussed.

Contributed Papers

2:40

4pAA6. Optimal loudspeaker placement for sound field reconstruction in geometrically constrained environments. Joshua Atkins and James West (Dept. of Elec. and Comput. Eng., Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21212)

The design of loudspeaker arrays for holographic sound reproduction systems involves sampling a given aperture by a set of loudspeakers. For full sphere apertures it can be proven that there are only five equidistant sampling methods (given by the five regular polyhedrons) which only allows for arrays with up to 20 elements. For larger arrays, non-equidistant design methods such as t-design and Gaussian sampling have been proposed. However, in realistic environments it is generally not possible to place loudspeakers on the floor or on walls at certain locations. Furthermore, it is not desirable to have the loudspeakers constrained to spherical radii. This work shows a method for loudspeaker array design that allows for consideration of these special cases. The method relies on the decomposition of the sound field into a spatially orthonormal basis of spherical harmonics. A heuristic method is used to minimize the eigenvalue spread of the spherical harmonic matrix. The loudspeaker weights are found by calculating the generalized inverse of this matrix. Constraints on the speaker positions, highest reproduction mode, and number of loudspeakers are imposed in the minimization. The results of a design for a 16 speaker array in our acoustics laboratory are discussed.

2:55

4pAA7. Acoustics in a home music listening room. Sergio Beristain (Acoust. Lab. E.S.I.M.E., P.O. Box 12-1022, Narvarte, 03020 Mexico Distrito Federal, Mexico, sberista@hotmail.com)

A room was designed with the sole purpose of listen all kinds of recorded music, with a high preference to classical and romantic music. It was decided to locate the room in the basement level of a large single family home, where the external noise control or the control of the sound going out were not the big issues, so that the internal acoustics became the main one. The acoustics was adjusted for clear and balanced sound, and the high-end audio system to be installed can handle monophonic, stereophonic, and surround signals, but stereo sources are paramount. The room is 4 m wide, 7 m long, and 2.8 m high, with the lower part of the walls used for the equipment and music source storage, while the floor, ceiling, and higher part of the walls are the places where the non-proprietary acoustic materials and the speakers are going to be located.

3:10

4pAA8. Acoustic calibration of the exterior effects room at the NASA Langley Research Center. Kenneth J. Faller, II, Stephen A. Rizzi, Jacob Klos (Structural Acoust. Branch, NASA Langley Res. Ctr., MS 463, Hampton, VA 23681, kenneth.j.faller@nasa.gov), William L. Chapin, Fahri Surucu (AuSIM, Inc., Mountain View, CA 94043), and Aric R. Aumann (Analytical Services and Mater., Inc., Hampton, VA 23666)

The exterior effects room (EER) at the NASA Langley Research Center is a 39-seat auditorium built for psychoacoustic studies of aircraft community noise. The original reproduction system employed monaural playback and hence lacked sound localization capability. In an effort to more closely recreate field test conditions, a significant upgrade was undertaken to allow simulation of a three-dimensional (3-D) audio and visual environment. The 3-D audio system consists of 27 full-range satellite speakers and four subwoofers, driven by a real-time audio server running a derivation of vector base amplitude panning. The audio server is part of a larger simulation system, which controls the audio and visual presentation of recorded and synthesized aircraft flyovers. The focus of this work is on the calibration of the 3-D audio system, including gains used in the amplitude panning algorithm, speaker equalization, and absolute gain control. Because the speakers are installed in an irregularly shaped room, the speaker equalization includes time delay and gain compensation due to different mounting distances from the focal point, filtering for color compensation due to different installations (half space, corner, and baffled/unbaffled), and crossover filtering.

Session 4pAB

Animal Bioacoustics, Signal Processing in Acoustics, and Noise: Topical Meeting on Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication II: Automated Classification of Animal Acoustic Signals II

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Sean K. Lehman, Cochair
Lawrence Livermore National Lab., Livermore, CA 94550

Chair's Introduction—1:10

Contributed Papers

1:15

4pAB1. Classification of marine mammal vocalizations using an automatic aural classifier. Paul C. Hines and Carolyn M. Ward (Defence R&D Canada, P.O. Box, 1012 Dartmouth, Nava ScotiaS B2Y 3Z7, Canada, paul.hines@drdcrddc.gc.ca)

Passive sonar systems are often used to detect marine mammal vocalizations in order to localize and track them. Unfortunately, transients generated by sources other than marine mammals can also trigger passive sonar systems which leads to a large number of false alarms. Furthermore, even in the case of a successful detection, classifying the genus is often required and this typically requires expertise not readily available on the vessel. Perceptual signal features—similar to those employed in the human auditory system—have been used to reduce false alarms in active sonar by automatically discriminating between target and clutter echoes. This contributes to improved sonar performance [Young and Hines, J. Acoust. Soc. Am. **122**, 1502–1517 (2007)]. Many of the features were inspired by research directed at discriminating the timbre of different musical instruments (a passive classification problem) which suggests it might be applied to classify marine mammal vocalizations. To test this hypothesis, the automatic aural classifier was trained and tested on a set of marine mammal vocalizations from a variety of species. This paper will provide an overview of the aural classifier's architecture, describe the preparation of the data set, including the attempt to provide ground-truth confirmation of the data, and present some preliminary results.

1:30

4pAB2. Processing burst pulse waveforms from odontocetes. Paul Hursky, Michael B. Porter (HLS Res. Inc., 3366 North Torrey Pines Court, Ste. 310, La Jolla, CA 92037), Tyler Olmstead, Marie Roch (San Diego State Univ., San Diego, CA 92182-7720), Sean Wiggins, and John Hildebrand (Univ. of California in San Diego, La Jolla, CA 92093)

A number of marine mammals produce an interesting vocalization called a burst pulse, which consists of a series of clicks whose repetition rate increases to the point that the entire waveform sounds more like a continuous buzz. These animals also produce individual clicks at a much less frequent rate, sometimes with a consistent inter-click interval, sometimes without. However, because these animals are often observed in groups that number in the hundreds at low rates, it is difficult to associate individual clicks with individual animals or to group multipath arrivals. A burst pulse waveform avoids these association ambiguities—the clicks making up a burst pulse are so frequent, and so consistent in amplitude and inter-click interval that they are easily associated with a single animal, and any multipath structure is consistently revealed. We will present results of processing a variety of burst pulses observed on array sensors from a series of experiments off the coast of California and discuss algorithms for detecting burst pulse waveforms and for extracting classification features from these waveforms.

1:45

4pAB3. Probabilistic bioacoustic signal extraction within spectrograms. T. Scott Brandes (Signal Innovations Group, Inc., 1009 Slater Rd., Ste. 200, Res. Triangle Park, NC 27703)

A methodology for probabilistic bioacoustic signal extraction within spectrograms of natural sound recordings is proposed. Probabilistic models of signal attributes are described for generating a host of likelihoods used to estimate the probability that individual pixels within a spectrogram represent part of a bioacoustic signal. The pixel probabilities result in a transformation of the spectrogram into a probability map of bioacoustic signal presence. It is shown that these probability maps create a dramatic increase in the bioacoustic signal-to-noise ratio within the spectrogram. These probability maps along with threshold filtering provide a means for image segmentation of the spectrogram, creating blocks of pixels that represent bioacoustic signals, facilitating feature and signal extraction. This methodology is applied to natural sound recordings of three quality types in a wide range of signal-to-noise ratios. In each instance, the probability mapping greatly increases the signal-to-noise ratio and, when applied as a threshold filter, is far more selective with pixel inclusion than threshold filtering applied based on sound level. Suggested applications include automated call recognition of birds, frogs, and insects from field recordings within a wide range of ambient noise.

2:00

4pAB4. Recognition of killer whale individuals from multiple stereotyped call types. Nicole Nichols, Les Atlas (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195, nmn3@u.washington.edu), Ann Bowles (Hubbs-SeaWorld Res. Inst., San Diego, CA 92109), and Marie Roch (San Diego State Univ., San Diego, CA 92182)

Recognition of marine mammal individuals will be an essential tool for obtaining demographic information of populations as habitat monitoring is increasingly being conducted with passive acoustics. Our previous research established the potential to identify killer whale individuals with passive acoustic recordings with up to 78% accuracy. The data for this result consisted of 75 vocalizations from four adult (two male, two female) killer whales of Icelandic origin in residence at SeaWorld San Diego. The stereotyped vocalizations used were all SD1a or SD1b, and caller identification was known to a high degree of certainty allowing for quantifiable validation. Classification was performed with 12 mel-frequency cepstral coefficients, and 12 delta and 12 delta-delta cepstral coefficients as inputs to a Gaussian mixture model. We initially chose to use one call type as the SD1 call was frequently produced by all four individuals and it eliminated potential ambiguity from identifying call type instead of individual. In the research we now present, we extend our prior methods to be used with multiple call types. Because individuals in the wild will not have a known frequency of call type usage, we propose a normalization method for training a classifier that is unbiased with respect to call-type frequency.

2:15

4pAB5. Autonomous underwater glider based embedded real-time marine mammal detection and classification. Tyler J. Olmstead, Marie A. Roch (Dept. of Comput. Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182, tyler.j.olmstead@gmail.com), Paul Hursky, Michael B. Porter (Heat, Light, and Sound Res. Inc., La Jolla, CA 92037), Holger Klinck, David K. Mellinger (Oregon State Univ., Newport, OR 97368), Tyler Helble, Sean S. Wiggins, Gerald L. D'Spain, and John A. Hildebrand (Univ. of California in San Diego, La Jolla, CA 92093)

Autonomous marine vehicles offer the potential to provide low-cost data suitable for passive acoustic monitoring applications of marine mammals. Due to their extremely low-power consumption and long range, gliders are an attractive option for long-term deployments. Challenges related to power availability, payload size, and weight have previously restricted the viability of marine mammal monitoring. As an example, the wide bandwidth of odontocete echolocation clicks requires a high sampling rate and poses challenges with respect to limitations in power, size, and weight of the deployed system. Recent developments in commercial off-the-shelf hardware driven by the mobile phone industry's need for multimedia-rich smart phones have resulted in low-power architectures capable of performing computationally demanding signal processing and stochastic recognition tasks in real time. We describe our work on a small form-factor, light-weight package used to perform real-time passive acoustic detection and classification. The system detects echolocation clicks using Teager energy. Echolocation clicks are then classified using cepstral features processed by a Gaussian mixture model. [Work sponsored by ONR.]

2:30

4pAB6. Analysis of cetacea vocalizations using ocean bottom seismic array observations, western Woodlark Basin, Papua New Guinea. Scott D. Frank (Dept. of Mathematics, Marist College, 3399 North Rd., Poughkeepsie, NY 12601, scott.frank@marist.edu) and Aaron N. Ferris (Weston Geophysical Corp., 181 Bedford. St., Lexington, MA 02420)

Over a 6-month period in 1999 an array of seven ocean bottom seismometers and seven bottom mounted hydrophones covered a 30-km^2 area at 2–3-km water depth in the western Woodlark Basin. The array's main task was the detection of microearthquakes associated with nearby active tectonics. However, it also made fortuitous broadband recordings of Cetacea vocalizations that we associate with finback whales. Each of these signals has a 10–15-s duration and they exhibit a repetitive pattern with 10–15-min period. The signals are well recorded across both the seismometers and hydrophones, which were sampled at 64 and 125 Hz, respectively. Time-varying spectral analysis demonstrates that the signals are frequency modulated in the 20–30-Hz band with most energy occurring at ~22 Hz. We employ array-based methods (e.g., optimal beam forming techniques) and acoustic transmission loss simulations to determine range and bearing of the acoustic sources. Preliminary analysis indicates that the acoustic sources originate near the array, and secondary signals may represent back-scattered energy from the short-wavelength, high-topographic features associated the active tectonics with in the basin.

2:45

4pAB7. Echolocation behavior of pairs of flying *Eptesicus fuscus* recorded with a Telemike microphone. Mary E. Bates (Dept. of Psych., Brown Univ., 89 Waterman St., Providence, RI 02912, mary_bates @brown.edu), Yu Watanabe, Yuto Furusawa, Emyo Fujioka, Shizuko Hiryu, Hiroshi Riquimaroux (Doshisha Univ., Kyotanabe, Japan), Jeffrey M. Knowles, and James A. Simmons (Brown Univ., Providence, RI 02912)

Four big brown bats (Eptesicus fuscus) were flown singly and in pairs in a room containing a sparse array of vertically hanging plastic chains as obstacles. Each bat carried a lightweight radio telemetry microphone (Telemike) that recorded their emitted echolocation sounds without artifacts from Doppler shifts, directional effects, and atmospheric attenuation. The broadcasts of both bats were also recorded with two stationary ultrasonic microphones located at the far end of the flight room. The echolocation broadcasts of bats flying singly were compared to those emitted when the bats were flown together. The principal change was shifting of harmonic frequencies very slightly (<5 kHz) away from each other and from frequencies used when flying alone. In contrast, the duration of emissions was more stable between single and double bat flights. Changes in ending frequency have been associated with a jamming avoidance response in big brown bats and could indicate attempts to avoid interference while flying with conspecifics in an enclosed space. [Work supported by ONR and NSF.]

3:00

4pAB8. Paradoxical reference frequency shifts during paired flights in Japanese horseshoe bats evaluated with a Telemike. Yuto Furusawa, Shizuko Hiryu, and Hiroshi Riquimaroux (Faculty of Life and Medical Sci., Doshisha Univ., 1-3 Miyakotani Tatara, Kyotanabe 610-0321, Japan)

In order to examine vocal behaviors for the echolocating bats to improve extraction of jammed echolocation signals, group flight experiments were conducted where two Japanese horseshoe bats were flown simultaneously in a flight chamber. Horseshoe bats are known to conduct Doppler-shift compensation, keeping their echo CF2 frequency constant to which their auditory system is sharply tuned. It was hypothesized that bats shifted their reference frequencies apart to avoid overlapping the reference frequencies. Echolocation sounds during sequential flights were recorded with an onboard wireless microphone, Telemike. Reference frequencies were investigated by measuring Doppler-shift compensated CF2 frequencies of the bats' own echoes. Recorded data show that the bats kept their echo CF2 frequency constant with a standard deviation (SD) of 50–80 Hz, indicating that they exhibited Doppler-shift compensation with the same accuracy as that used for flying alone (SD = 83.2 Hz). However, the bats did not shift their reference frequency toward increasing the frequency difference even when the initial difference in the reference frequencies between individuals was less than 100 Hz. Rather, the bats tended to make their reference frequencies closer during group flight compared to those for single flights. These findings suggest that multiple bats may squeeze their echo CF2 frequencies into a narrow frequency range. In addition, horseshoe bats may adapt to acoustic interference without expanding the frequency difference between individuals. [Work supported by ONR grant.]

3:15

4pAB9. Miniaturization of insect-inspired acoustic sensors. Erick Ogam (Laboratoire de Mcanique et d'Acoustique, CNRS, UPR7051, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France), Franck Ruffier (Institut des Sci. du Mouvement, 13009 Marseille France), Armand Wirgin (CNRS, 13402 Marseille Cedex 20, France), and Andrew Oduor (Maseno Univ., Maseno, Kenya)

Insects can navigate and follow sound sources with precision in complex environments using efficient principles based on particular and appropriate sensory-motor processing. It is precisely these principles, little understood at this point, that we want to understand, model, and validate by rebuilding them on a micro-flying device. To achieve this goal, a new bio-mimetic acoustic sensor inspired by the cricket's auditory system has been developed and tested in an anechoic chamber. The performance of the processing algorithm, the aperture of the auditory system, and interferences caused by the geometry of the sensor itself were first evaluated: synthetic signals were generated by a boundary element model taking into account the three-dimensional geometry of the sensor to characterize its performance in free space. A processing based on cross-correlation for localizing and tracking an acoustic source was also studied to assess whether the observed limitations were inherent to the geometry of the sensor or due to limitations of the cricket inspired processing algorithm. The sensor worked well with several synthetic chirped cricket songs. [Work supported by CNRS PIR Neuroinformatique grant SonoBot.]

3:30—3:45 Break

3:45—4:30 Demonstrations and Discussions

Session 4pAO

Acoustical Oceanography and Underwater Acoustics: Impact of Shallow Water Acoustic Propagation by Linear Internal Waves and Neutrally Buoyant Intrusions

Dajun Tang, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105

David L. Bradley, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804–0030

Chair's Introduction—1:00

Invited Papers

1:05

4pAO1. Three-dimensional finestructure and vertical microstructure in thermohaline intrusions. Michael C. Gregg (522 Henderson Hall, Univ. Washington, 1013 NE 40th St., Seattle, WA 98105), Andrei Y. Shcherbina, Matthew H. Alford, and Ramsey R. Harcourt (Univ. Washington, Seattle, WA 98105)

Thermohaline intrusions have intrigued physical oceanographers since the first XBT section was published by Spilhaus in 1939. Characterized by vertical gradients of temperature and salinity having compensating density effects, intrusions are meters to tens of meters thick and extend laterally for hundreds of meters to many kilometers. Boundaries often have TS staircases produced by salt fingering or diffusive layering. Large diapycnal fluxes have been inferred from sections showing intrusions sloping across isopycnals, but these inferences are questionable, as isopycnal movements can give the illusion of diapycnal motion. To address these issues, in 2007 we used a depth-cycling towed body, a neutrally buoyant Lagrangian float, and microstructure profilers to map and follow intrusions in the subtropical front north of Hawaii. All intrusions were connected to the front, i.e., none was isolated lenses, some as long, thin folds, and others resembled tongues. About 10 m thick, the sheets protruded 10 km into the warm, saline side of the front, were coherent laterally for 10–30 km, and lasted at least 1 week, evolving at both inertial (23 h) and sub-inertial (~10 day) time scales. They appeared to be formed by advective deformation of the front driven by sub-mesoscale two-dimensional motions.

1:25

4pAO2. Internal waves, intrusions, and coastal zone acoustics. Timothy F. Duda, Andone C. Lavery (Dept. AOPE, MS 11, Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, tduda@whoi.edu), and Jon M. Collis (Colorado School of Mines, Golden, CO 80401)

Ubiquitous linear internal waves are known to dominate fluctuations of long-range low-frequency sound in the deep ocean. In the coastal zone, however, variations in low-frequency-sound propagation caused by large episodic nonlinear internal waves are often dominant and have been studied in detail. Apart from this, sound-speed anomalies from smaller but pervasive linear internal waves can also influence propagation, as do density-compensated temperature/salinity intrusions found in areas near fronts. The physical nature of each of these two phenomena and existing models of their effects on propagation will be reviewed. The possibility of using high-frequency acoustics to remotely examine and map intrusions via double-diffusive microstructure scattering will also be examined. The anisotropy of the coastal internal wave field will be discussed, and subsequent acoustic propagation anisotropy will be examined using data from the multi-PI Shallow-Water 2006 experiment in the Mid-Atlantic Bight. [Research supported by the Office of Naval Research.]

Contributed Papers

1:45

4pAO3. Analysis of horizontal coherence during the transverse acoustic variability experiment. Peter C. Mignerey and David J. Goldstein (Acoust. Div. 7120, Naval Res. Lab., Washington, DC 20375)

The transverse acoustic variability experiment took place in the northern limit of the East China Sea in 65–80 m of water. 300-Hz and 500-Hz cw acoustic signals were transmitted over distances of 33 and 20 km to a broadside HLA where they were received at 22–27 dB SNR. Fourteen estimates of transverse mutual-coherence functions show high correlation with scale lengths on the order of 1000–1700 m. A towed CTD chain provided simultaneous measurements of the sound-speed fluctuation spectrum due to internal waves in an effort to determine whether the observed shallow-water coherence is understandable within the framework of a normal-mode formulation of path-integral theory. An important property of the mutual-coherence function is its behavior at small hydrophone separations. Here the phase-structure function is expected to follow a power law. Measured logarithmic phase-structure function have slopes that vary between 0.6 and 1.1, which disagree with path-integral solutions that predict slope 2. The discrepancy between observation and data is not currently understood. [Work supported by the Office of Naval Research.]

2:00

4pAO4. Performance of the transverse acoustic variability experiment horizontal line array. David J. Goldstein and Peter C. Mignerey (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, david.goldstein@nrl.navy.mil)

The transverse acoustic variability experiment was designed to study the impact of shallow-water environmental fluctuations on acoustic coherence in the horizontal plane, transverse to the direction of propagation. The experiment was conducted during August 2008 in the East China Sea. A criti-

cal component to the success of the experiment was a two-segment 96-channel horizontal line array, moored on the sea floor at an approximate depth of 80 m. An analysis of the performance of this array is presented, including pulse arrival time statistics, beam patterns, gains, and signal-to-noise ratio. Methods used to determine the relative positions of array hydrophones are also discussed. [Work supported by the Office of Naval Research. We thank the staff at KORDI and the crews aboard the R/V Eardo and Sunjin for their assistance and bountiful hospitality.]

2:15

4pAO5. Mode-2 internal waves impact on acoustic signal properties. Marshall Orr (Acoust. Div., The Naval Res. Lab., 4555 Overlook Ave., SW, Washington, DC 20375)

The properties of mode-2 internal waves repeatedly observed on the New Jersey Shelf (USA) with high-frequency acoustic flow visualization systems will be overviewed. The impact of the mode-2 internal waves on the range/depth distribution of acoustic signal properties will be discussed.

2:30

4pAO6. Effects of anisotropic internal waves on acoustic propagation within the East China Sea. Chad M. Smith (cms561@psu.edu) and David L. Bradley (Graduate Program in Acoust., Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804-0030)

Using environmental and acoustic data from the TAVEX 2008 Korean /U.S. cooperative experiment, the anisotropic spatial properties of internal waves recorded during CTD tows and their connection to temporal variations in recorded acoustic data from a horizontal line array will be discussed. Structure of internal waves, wave travel speed and trajectory, as well as computational modeling of the acoustic field and its comparison with recorded data will also be covered. [Work supported by the Office of Naval Research Contract No. N00014-08-1-0455.]

2:45

4pAO7. Adjoint study of water column variability and propagation in the transverse acoustic variability experiment (TAVEX 2008). Michelle M. Kingsland and David L. Bradley (Appl. Res. Lab., Graduate Program in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16801, mbm224@psu.edu)

Internal wave and other water column activity in the conductivity, temperature, and depth (CTD) data collected in the East China Sea as a part of TAVEX 2008 are used as a basis for analyzing acoustic propagation at 500 Hz between a fixed source and fixed horizontal line array placed on the seafloor. Because the CTD data were not taken coincident with the acoustic propagation path (in general), adjoint modeling methods will be used to study acoustic data and prediction misfit. Usefulness of the adjoint method for these experimental conditions will be discussed. [Work supported by ONR.]

3:00

4pAO8. Periodicity of intensity fluctuations in shallow water in the presence of solitary internal waves. T. C. Yang (Naval Res. Lab., Washington, DC 20375, tc.yang@nrl.navy.mil)

This paper investigates the intensity variation with time of a broadband signal in the presence of moving solitary internal waves. The autocorrelation of the received impulse responses summed over a vertical line array of receivers (the so called Q function) is measured from the SWARM95 data using broadband pulses received on the NRL array at a range ~42 km from the source. The signals are m-sequences centered at 400 Hz with a bandwidth of 100 Hz. The peak of the Q function shows a periodicity of 0.042 cycles/min, with a value changing with time depending on the position of the internal wave packets. Coupled mode equation is used to calculate mode coupling induced by the solitary internal waves and explain the observed periodicity of the intensity fluctuations. The results suggest that the positions and some parameters of the solitary waves could be remotely estimated from underwater sensors. [Work supported by the Office of Naval Research.]

3:15—3:30 Break

Invited Papers

3:30

4pAO9. Observed space-time scales of internal waves and finescale intrusions on the New Jersey Shelf and their likely acoustic implications. John Colosi (Naval Postgrad. School, Monterey, CA 93943), Tim Duda, Ying-Tsong Lin, Jim Lynch, and Arthur Newhall (Woods Hole Oceanograph. Inst., Woods Hole MA 02543)

It has been known for some time that ocean density surfaces possess multi-scale variations in temperature (T) and salinity (S), termed spice (hot and salty, cold and fresh) by Munk in 1981. Compensating T and S variations along a surface of constant density have reinforcing sound speed anomalies and thus have important acoustic implications. With the advent of new technologies to observe small scale, rapidly changing ocean thermohaline structure, spice variations, and their acoustic effects are just starting to be understood. Better known are the impacts of internal waves which vertically advect ocean sound speed structure. This talk will present analysis of moored T and S observations on the New Jersey Shelf region during the SW06 experiment in which the sound speed effects of internal waves and spice can be approximately separated. These results will be contrasted with similar measurements made in the deep waters of the Philippine Sea. Using recent results from coupled mode theory, some discussion of the acoustic implications of the internal wave and spice structure will be presented. In particular, the phase randomizing effects of linear internal waves and spice will diminish the mean acoustical influence of intense but localized nonlinear internal waves.

3:50

4pAO10. Characterizing water column variability and its impact on underwater acoustic propagation. Kyle M. Becker (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-0030) and Megan S. Ballard (Appl. Res. Labs., Univ. Texas at Austin, 10000 Burnet Rd., Austin, TX 78758)

During the Shallow Water Experiment 2006 (SW06), a low-frequency acoustic source was towed out and back along radials from a receiver array located at the origin. Measurements were made along three radials, separated by 45 deg, with each track repeatedly traversed over a period of 6–10 h. Concurrently with the acoustic measurements, the water column sound speed field was estimated both along the source track and at the receiver location. The objective was to fully characterize the environment in the water column over the acoustic path during transmissions. In this way, variability in the acoustic field measured along the repeated tracks could be associated with variability in the water column. The variability in the water column, measured independently at the source and at the receiver, is

examined and related to variability observed in the acoustic fields. Spatial and temporal variability is characterized on transect-to-transect, radial-to-radial, and day-to-day basis. Over 3 days, variability in the water column could be attributed to thermohaline intrusions, linear and non-linear internal waves, as well as mesoscale variability. This talk will focus on the requirements for characterizing these features and understanding their impact on acoustic propagation and prediction. [Work supported by ONR.]

Contributed Papers

4:10

4pAO11. Reverberation clutter from combined internal wave refraction and bottom backscatter. Dajun Tang and Frank Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, djtang@apl.washington.edu)

A hypothesis is proposed that one mechanism for clutter observed in shallow water reverberation measurements is due to the combined effect of forward scatter and subsequent backscatter. The forward scatter refracts part of the sound from low grazing angle to high grazing angle, then being backscattered by bottom roughness. The refracted sound impinges onto the bottom at higher grazing angle, resulting in higher backscatter because of elevated bottom scattering cross section as compared to that at lower grazing angles. The resultant reverberation will stand out as a target-like clutter. An example is presented as a non-linear internal wave propagates in a shallow water channel, resembling conditions found on the New Jersey Shelf. The effect on reverberation of the internal wave as clutter is investigated using a time-domain numerical model. The model uses parabolic equation for the two-way propagation and first order perturbation approximation for bottom backscatter. [Work supported by ONR.]

4:25

4pAO12. Acoustic propagation effects from saline intrusions across the shelfbreak front off the New Jersey coast in summer. Arthur E. Newhall, Ying-Tsong Lin, James F. Lynch, Timothy F. Duda, Glen G. Gawarkiewicz (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

In the summer of 2006, the multi-task, interdisciplinary Shallow Water 2006 (SW06) experiment was conducted on the continental shelf and shelfbreak regions off the New Jersey coast. During SW06 the shelfbreak front variability was clearly characterized by combined measurements of conductivity-temperature-depth casts, long-period (5 weeks) and quick sampling oceanographic sensor moorings, and surveys with a Scanfish, a ship-towed undulating vehicle carrying multiple sensors. This data set indicates that the slope water penetrated onto the shelf via neutrally buoyant intrusions. Due to the higher temperature and salinity, these frontal intrusions, seen as thin layers near the seasonal pycnocline, increased the local sound speed and accounted for variations in acoustic propagation. In this paper, we will use field data and PE numerical acoustic simulations to study the impact of these frontal intrusions. Specific examples on the creation of vertical acoustic double ducting and horizontal ducting formed by these intrusions and their modulation by linear internal waves are discussed. [Work is supported by the Office of Naval Research.]

4:40

4pAO13. Propagation of low-frequency sound through densely sampled internal waves. Megan S. Ballard (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758, meganb@arlut.utexas.edu) and Kyle M. Becker (Appl. Res. Lab., Penn State, State College, PA 16804)

It is well known that nonlinear internal gravity waves can have a significant effect on acoustic propagation. In particular, horizontal refraction and ducting of sound can occur when the acoustic propagation path is aligned with fronts of internal waves. Additionally, more complicated effects are associated with internal waves that are curved and/or truncated. In this talk, observations of internal waves measured on the New Jersey shelf area of the north Atlantic and their effects on the acoustic field are presented. During the experiment, a low-frequency source broadcasting continuous tone was towed repeatedly out and back along radials with respect to a VLA. The ship's track was oriented parallel to the shelf break so that the acoustic propagation path was roughly aligned with internal waves propagating up the shelf. Internal waves were measured at the location of the receivers by temperature sensors on the VLA, at the location of the source by a towed CTD chain, as well as by a cluster of 16 environmental moorings located adjacent to experiment site. The high-spatial sampling of the environment allows for identification of range-varying features of internal waves.

4:55

4pAO14. Horizontal reflection of a low-frequency sound signal from a moving nonlinear internal wave front. M. Badiey (College of Earth, Ocean, and Environment, Univ. of Delaware, Newark, DE 19716), J. F. Lynch, Y.-T. Lin (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and B. G. Katsnelson (Voronezh Univ., Voronezh 394006, Russia)

Simultaneous measurements of acoustical and internal waves are reported while an internal wave approaches an acoustic track during the SW06 experiment. The incoming internal wave packet acts as a moving layer reflecting and refracting acoustic waves in the horizontal plane. The mechanism of this interaction is shown in received acoustic data on a vertical hydrophone array using a modal approach. It is shown that the wave front of internal waves behaves as selective filter, depending on the mode number and frequency of the broadband signal. In other words the reflection coefficient in horizontal plane depends on mode number and frequency. Experimental data analysis shows good agreement with the theory. [Work supported by ONR 321OA.]

5:10

4pAO15. Acoustic wave propagation and scattering in turbulence and internal waves. Tokuo Yamamoto N (Div. of Appled Marine Phys., RSMAS, Univ. of Miami, Miami, FL 33149)

Acoustic wave propagation and scattering in turbulence and internal waves in Kanmon Strait and the outer New Jersey shelf are measured in a bistatic source-receiver configuration at frequency 5500 Hz. While internal wave dominates on the open New Jersey shelf, the turbulence dominates in the Kanmon Strait channel. Initially strong internal wave scattering is reduced rapidly due to strong mixing in the shallow strait. On the flat and open continental shelf, internal wave and warm water intrusion dominate while turbulence is near absent. The parabolic equation code predictions agree with data very well. The bistatic scattering data are processed for bistatic angles and the effects of Doppler shift on the mean flow and the broadening of peek spectra on the turbulent fluctuations. The mean flow and turbulence fluctuation obtained from the Doppler shift agreed with the two parameters obtained by the reciprocal acoustic transmission.

5:25

4pAO16. Using ocean ambient noise cross-correlations for noise-based acoustic tomography. Charlotte Leroy, Karim G. Sabra (School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332-0405), Philippe Roux (LGIT, Grenoble, France), William Kuperman, William Hodgkiss, and Shane Walker (UC San Diego, La Jolla, CA)

Recent theoretical and experimental studies have demonstrated that an estimate of the Green's function between two hydrophones can be extracted passively from the cross-correlation of ambient noise recorded at these two points. Hence monitoring the temporal evolution of these estimated Green's functions can provide a means for noise-based acoustic tomography using a distributed sensor network. However, obtaining unbiased Green's function estimates requires a sufficiently spatially and temporally diffuse ambient noise field. Broadband ambient noise ([200 Hz–20 kHz]) was recorded continuously for 3 days during the SWAMSI09 experiment (next to Panama City, FL) using two moored vertical line arrays (VLAs) spanning the 13-m water column and separated by 150 m. The feasibility of noise-based acous-

tic tomography was assessed in this dynamic coastal environment over the whole recording period. Furthermore, coherent array processing of the computed ocean noise cross-correlations between all pairwise combinations of hydrophones was used to separate acoustic variations between the VLAs caused by genuine environmental fluctuations–such as internal waves–from the apparent variations in the same coherent arrivals caused when the ambient noise field becomes strongly directional, e.g., due to an isolated ship passing in the vicinity of the VLAs.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Ultrasonically Activated Agents

Oliver Kripfgans, Chair
Univ. of Michigan, Radiology Dept., Zina Pitcher Pl., Ann Arbor, MN 48109-0553

Chair's Introduction—12:55

Invited Papers

1:00

4pBB1. Cavitation and vaporization threshold of perfluorocarbon droplets. Kelly Schad and Kullervo Hynynen (Dept. Medical Biophys., Univ. Toronto, 2075 Bayview Ave., Toronto, ON, Canada)

Focused ultrasound therapy can be enhanced with microbubbles by thermal and cavitation effects. However, localization of treatment becomes difficult as bioeffects can occur outside of the target region. Spatial control of gas bubbles can be achieved with acoustic vaporization of perfluorocarbon droplets. This study was undertaken to determine the acoustic parameters for bubble production by droplet vaporization and how it depends on the acoustic conditions and droplet physical parameters. Droplets of varying sizes were sonicated *in vitro* with a focused transducer and varying frequencies and exposures. Simultaneous measurements of the vaporization and inertial cavitation thresholds were performed. The results show that droplets cannot be vaporized at low frequency without inertial cavitation occurring. However, the vaporization threshold decreased with increasing frequency, exposure, and droplet size. In summary, we have demonstrated that droplet vaporization is feasible for clinically relevant sized droplets and acoustic exposures.

1:20

4pBB2. Applications of acoustic droplet vaporization in diagnostic and therapeutic ultrasound. Mario L. Fabiilli, Oliver D. Kripfgans, Man Zhang, Kevin J. Haworth, Andrea H. Lo, Paul L. Carson, and J. Brian Fowlkes (Dept. of Radiology, Univ. of Michigan, 3225 Med. Sci. I, 1301 Catherine St., Ann Arbor, MI 48109, mfabiill@umich.edu)

Micron- and nano-sized colloids are being studied in diagnostic and therapeutic applications of ultrasound (US). Unlike clinically utilized microbubbles, emulsions possess unique physiochemical properties that could translate into distinct, clinical benefits beyond conventional contrast agents. Droplets, composed of a superheated liquid, can be phase transitioned into bubbles using US, a process known as acoustic droplet vaporization (ADV). Droplets, transpulmonary in size, transition into bubbles upon ADV and can reach diameters that occlude capillaries and arrest blood flow in the vascular bed. Examples of ADV in diagnostic and therapeutic applications will be presented. First, ADV has been used in phase aberration correction in transcranial US imaging. Second, ADV-generated microbubbles can reduce and occlude renal perfusion *in vivo*. Third, the effects of thermal therapy using high-intensity focused ultrasound (HIFU) has been enhanced and controlled more effectively using ADV. Fourth, ADV has been utilized as a release mechanism for therapeutic agents that are incorporated into the emulsion. In all applications, the physiochemical properties of the droplets coupled with the spatial and temporal control afforded by ADV-generated microbubbles are crucial to the success of each ADV development. [This work was supported in part by NIH Grant 5R01EB000281.]

1:40

4pBB3. Dodecaperfluoropentane emulsion for oxygen delivery. Evan Unger, Terry Matsunaga, Russel Witte, Ragnar Olafsson (Dept. of Radiology, Univ. of Arizona, 1501 N. Campbell Ave., Tucson, AZ 85724), Arthur Kerschen, Melissa Dolezal, and Jenny Johnson (NuvOx Pharma, LLC, 1635 E. 18th St., Tucson, AZ 85719)

Dodecaperfluoropentane emulsion (DDFPe) was prepared from DDFP and surfactant PEG-Telomer-B and studied for particle sizing, storage stability, acoustically *in vitro* at 37 °C, and for oxygen transport in porcine model of hemorrhagic shock and as radiation sensitizer in murine model of tumor hypoxia. DDFPe had mean particle size of about 200 nm stable on storage at room temperature for over 2 years. Compared to microbubbles of perfluorobutane (PFB), more than 100-fold higher DDFPe was needed for comparable acoustic backscatter compared to PFB. Acoustic reflectivity increased with higher MI ultrasound but was still much less for DDFPe than for PFB. DDFPe was highly effective oxygen therapeutic in models of hemorrhagic shock and tumor hypoxia. A dose of as little as 0.6 cc/kg of DDFPe was sufficient to prevent death from hemorrhagic shock. In hypoxic tumors a similar dose reversed tumor hypoxia and restored sensitivity to radiation therapy. No ultrasound was used in these experiments for *in vivo* oxygen delivery. Compared to liquid perfluorocarbons which have been studied as oxygen therapeutics, DDFPe uses a much lower dose, about 1%. DDFPe is a promising oxygen transport therapeutic. Additional work is underway to try to develop DDFPe for this application.

2:00

4pBB4. Phase change nanodroplet for sensitizer for thermal and cavitational effects of ultrasound. Ken-ichi Kawabata, Rei Asami, Hideki Yoshikawa, Takashi Azuma (Life Sci. Res. Ctr., Central Res. Lab., Hitachi, Ltd., Higashi-Koigakubo, Kokubunji, Tokyo 185-8601, Japan, kenichi.kawabata.ap@hitachi.com), and Shin-ichiro Umemura (Tohoku Univ., Aramaki, Aoba-ku, Sendai 980-8579, Japan)

Although HIFU therapy is very low invasive method for tumor treatment, its time inefficiency and the absence of focus detecting mechanism of current systems limit the number of applicable cases. For an improved HIFU tumor treatment system, we propose to utilize a non-echogenic liquid nanodroplet which turns into highly echogenic microbubbles upon non-therapeutic ultrasound pulses. Such a nanodroplet would give echographic information on the focus of HIFU when exposed to the pulse and accelerate HIFU treatment by enhancing the ultrasonic energy deposition as microbubble. As such a nanoparticle system, we developed a phase change nanodroplet (PCND) which consists of superheated perfluorocarbon nanodroplet coated with PEGylated phospholipids. Significance of PCND to enhance ultrasonically induced therapeutic effects was investigated on mice, rats, and rabbits. At frequencies of 2 and 3 MHz, it was found that the presence of PCND (and generated microbubbbles) halved the intensity threshold for inducing thermal damage. At frequency around 1 MHz, it was found that the PCND halved the threshold for inducing damage in murine tumor, and the damage was induced both by thermal and cavitational effects of ultrasound. [This work was in part entrusted by the New Energy and Industrial Technology Development Organization, Japan.]

2:20

4pBB5. Relationship between cavitation, rapid loss of echogenicity, and drug release from echogenic liposomes. Kirthi Radhakrishnan (Dept. of Biomedical Engg., Univ. of Cincinnati, 231 Albert Sabin Way, MSB 6155, Cincinnati, OH 45267, radhakki@mail.uc.edu), Jonathan A. Kopechek, Kevin J. Haworth (Univ. of Cincinnati, Cincinnati, OH), Shaoling Huang, David D. McPherson (Univ. of Texas Health Sci. Ctr., Houston, TX), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH)

Echogenic liposomes (ELIPs), encapsulating air and drug, are being developed for use as vesicles for ultrasound-mediated drug release. Both calcein and a thrombolytic drug have been encapsulated in ELIP and released with 6-MHz color Doppler ultrasound in an *in vitro* flow system. To elucidate the role of stable or inertial cavitation in loss of echogenicity and drug release, ELIPs were insonified with 6-MHz pulsed Doppler ultrasound within the same flow system. A 10-MHz focused passive cavitation detector (PCD) was placed confocally with the pulsed Doppler sample volume. B-mode images and PCD signals were recorded during ELIP infusion (5 ml/min) and exposed to pulsed Doppler over a range of peak rarefactional pressure amplitudes (0–1.4 MPa) at two pulse repetition frequencies (1250 and 2500 Hz). The mean digital intensities were calculated on each image within two regions of interest placed upstream and downstream of the Doppler sample volume. The thresholds for rapid loss of echogenicity were ascertained from six B-mode images acquired over 30 s. These thresholds were compared to stable and inertial cavitation thresholds. The implications of these data on the mechanism of rapid liberation of gas and drug release will be discussed. [Work supported by the NIH R01 HL059586 and NIH R01 HL7400]

2:40

4pBB6. Bubble evolution in acoustic droplet vaporization. Adnan Qamar, Z. Z. Wong (Dept. of Biomedical Eng., Univ. of Michigan, 1101 Beal Ave., Ann Arbor, MI 48109), J. Brian Fowlkes, and Joseph L. Bull (Univ. of Michigan, Ann Arbor, MI 48109, joebull@umich.edu)

We present an overview of our recent theoretical and computational work on acoustic droplet vaporization within a confined tube and compare to our experimental data. This work is motivated by a developmental gas embolotherapy technique that involves injecting superheated transvascular liquid droplets and subsequently vaporizing the droplets with ultrasound to selectively form vascular microbubbles. The theoretical model describes the rapid phase transition from a superheated dodecafluoropentane (DDFP) droplet to the vapor phase via nucleation within the DDFP droplet. The theoretical results are compared to results from our computational models and high-speed camera experiments, and close agreement between the experimental bubble evolution and the prediction of the model is noted. Even though we consider bubbles that are small compared to the tube diameter, it is demonstrated that the tube affects the bubble evolution and resulting flow fields. [This work is supported by NIH Grant R01EB006476.]

3:00—3:30 Break

Contributed Papers

3:30

4pBB7. Ultrasound-mediated tumor chemotherapy with drug-loaded phase-shift nanoemulsions. Natalya Rapoport, Kweon-Ho Nam, Douglas A. Christensen, and Anne M. Kennedy (Dept. of Bioengineering, Univ. of Utah, 72 S. Central Campus Dr., Rm. 2646, Salt Lake City, UT 84112, natasha.rapoport@utah.edu)

This paper describes droplet-to-bubble transition in block copolymer stabilized perfluoropentane nanoemulsions used for ultrasound-mediated tumor chemotherapy. Three physical factors that trigger droplet-to-bubble transition in liquid emulsions and gels have been evaluated, namely, heat, ultrasound, and injections through fine-gauge needles. Among those listed, ultrasound irradiation was found the most efficient factor. Tumor accumulation of nanodroplets after systemic injections to tumor bearing mice was confirmed by ultrasound imaging. Efficient ultrasound-triggered drug release from tumor-accumulated nanodroplets was confirmed by dramatic regression of ovarian, breast, and orthotopic pancreatic tumors treated by systemic injections of drug-loaded nanoemulsions combined with tumor-directed therapeutic ultrasound. No therapeutic effect from the nanodroplet /ultrasound combination was observed without the drug, indicating that therapeutic effect was caused by the ultrasound-enhanced chemotherapeutic action of the tumor-targeted drug, rather than the mechanical or thermal action of ultrasound itself. The mechanism of drug release in the process ultrasonically activated droplet-to-bubble conversion is discussed. [The work has been supported by the NIH R01 EB1033 grant to N.R.]

3:45

4pBB8. Passive mapping of cavitation activity for monitoring of drug delivery. Costas D. Arvanitis, Miklos Gyongy, Miriam Bazan-Peregrino, Bassel Rifai (Dept. of Eng. Sci., Inst. of Biomedical Eng., Univ. of Oxford, Old Rd. Campus Res. Bldg., Headington, Oxford OX3 7DQ, United Kingdom), Leonard W. Seymour, and Constantin C. Coussios (Univ. of Oxford, Oxford OX3 7DQ, United Kingdom)

Acoustic cavitation has been recently shown to have tremendous potential as a mechanism for releasing and enabling improved uptake and biodistribution of therapeutic agents for the pharmacological treatment of cancer, cardiovascular, and neurodengenerative diseases. Passive cavitation mapping makes it possible to localize, characterize, and quantify cavitation activity by detecting the acoustic emissions produced by cavitating bubbles on an array of receivers, which are then back-propagated to form a cavitation-selective image with high spatio-temporal accuracy. In the present study, a 64-element diagnostic linear array was used to construct passive cavitation maps during pulsed HIFU exposure of a narrow channel transporting fluorescent dextran nanoparticles through a biocompatible porous hydrogel in the presence or absence of ultrasound contrast agents. Spectral analysis of the resulting acoustic emissions confirmed that different cavitation regimes were successfully instigated over the HIFU parameter range that was investigated. Superposition of cavitation selective images over the treatment course was used to construct maps of cavitation activity and dose as a function of location for treatment assessment and was found to correlate well with dextran extravasation. It is concluded that passive mapping of cavitation activity shows great promise for assessment and real-time monitoring of cavitation based drug delivery.

4:00

4pBB9. Combined ultrasonic and video-microscopic characterization of ultrasound contrast agents in a flow phantom. Parag V. Chitnis, Paul Lee, Jonathan Mamou (F. L. Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., 9th Fl., New York, NY 10038), Paul A. Dayton (UNC-NCSU, Chapel Hill, NC 27599), and Jeffrey A. Ketterling (Riverside Res. Inst., New York, NY 10038)

A high-frequency ultrasound pulse-echo system has been combined with a video microscope system for characterizing the response of two types of ultrasound contrast agents (UCAs). A focused-flow phantom was constructed using a glass micropipette (inner diameter 30 μm) mounted vertically in a test tank. The UCA solution was pumped using a syringe pump at a constant flow rate of 40 μl/min. A high-speed CCD camera operating at 10 000 frames/s with an exposure time of 4 μs was used to capture backlit images of individual UCAs. The camera was equipped with a 100× water-immersion objective with a field-of-view of 70×70 μm. A 40-MHz, focused transducer was aligned at the micropipette and used in pulse-echo mode to sonicate individual UCAs and receive the corresponding radio-frequency backscatter signals. The UCAs were subjected to ultrasound at varying incident-pressure levels (0.8–3 MPa) and pulse durations (10, 15, and 20 cycles). Echo signals from individual UCAs were windowed and spectra were calculated. The magnitude of the fundamental (40-MHz) backscatter was observed to increase monotonically with UCA diameter. Correlation between the magnitude of the sub-harmonic (20-MHz) response of UCAs and their size was investigated. [Work supported by NIH EB006372.]

4:15

4pBB10. Dynamic response of bubbles within a compliant tube. Neo Jang, Aaron Zakrzewski, Christopher Jensen, Robert Halm (Dept. of Mech. Eng., Univ. of Rochester, Rochester, NY 14627), and Sheryl M. Gracewski (Univ. of Rochester, Rochester, NY 14627, grace@me.rochester.edu)

The dynamic response of bubbles in a liquid that are partially constrained by a surrounding tube or channel is important in a variety of fields, including diagnostic and therapeutic biomedical ultrasound and for microfluidic devices. In this study, numerical simulations, lumped parameter models, and experiments are used to investigate the effects of a surrounding tube on a bubble's response to acoustic excitation. In particular, a coupled boundary element and finite element model and COMSOL MULTIPHYSICS models have been developed and used to investigate the nonlinear interactions of this three-phase system. Simulation results were compared to experimental measurements obtained using a scaled balloon model. The effects of tube parameters and bubble interactions on a bubble's natural frequency important for proposed clinical applications of ultrasound are investigated. Resonance frequencies agree well with one-dimensional lumped parameter model predictions for a bubble well within a rigid tube, but deviate for a bubble near the tube end. Simulations also predict bubble translation along the tube axis and the aspherical oscillations and induced tube stresses at higher amplitudes. [Work supported by NIH and NSF CMMI.]

4:30

4pBB11. Using optically activated nanoparticles to promote controlled lesion formation from high-intensity focused ultrasound exposures. James R. McLaughlan (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, jmcl@bu.edu), Todd W. Murray (Univ. of Colorado, Boulder, Boulder, CO 80309), and Ronald A. Roy (Boston Univ., Boston, MA 02215)

The absorption of laser light in tissue can be enhanced through the use of gold nanoparticles. This enhancement can improve the signal-to-noise ratio of the thermoelastic emissions used for photoacoustic tomography (PAT) and photoacoustic microscopy (PAM). The ability to functionalize nanoparticles allows them to be used for the selective targeting and destruction of cancer cells through the formation of vapor bubbles. However, the laser fluence required to generate vapor bubbles from nanoparticles typically exceeds the maximum permissible exposure for *in-vivo* applications. In a previous study, it was found that the combination of laser light with high-intensity focused ultrasound (HIFU) significantly reduced the fluence and pressure thresholds for bubble formation. The presence of bubbles, or specifically inertial cavitation, in HIFU exposures is believed to locally enhance the heating, resulting in effective tissue ablation at lower HIFU intensities. Nanoparticle-doped tissue phantoms (polyacrylamide gels containing bovine serum albumin) were exposed to a continuous wave (5–30-s) 1.1-MHz HIFU field with and without pulsed laser light (532 nm, 1–10 Hz) to assess the potential for HIFU lesion enhancement from the controlled and repeatable generation of inertial cavitation. [Work supported by the Gordon Center for Subsurface Sensing and Imaging Systems, NSF ERC Award No. EEC-9986821.]

4:45

4pBB12. Effects of vaporized phase-shift nanoemulsion on high-intensity focused ultrasound-mediated heating and lesion formation in gel phantom. Peng Zhang and Tyrone Porter (Dept. of Mech. Eng., Boston Univ., 110 Cummington St, Boston, MA 02215, tmp@bu.edu)

In this study, we demonstrate the feasibility of using a phase-shift nanoemulsion (PSNE) as nuclei for bubble-enhanced heating and lesion formation during high-intensity focused ultrasound (HIFU) thermal ablation. In our experiments, different concentrations of PSNE (5%–15%) uniformly distributed throughout gel phantoms made from acrylamide and albumin were sonicated with a 1.1 MHz HIFU transducer. The PNSE at the transducer focus were first vaporized into bubbles and then acoustically driven with continuous wave HIFU at different amplitudes. Bubble-enhanced heating resulted in albumin denaturation and lesion formation. HIFU-mediated heating was conducted with and without vaporized PSNE, and the lesion size and location relative to the transducer focus were measured after HIFU exposure. In the presence of vaporized PSNE, lesions were formed more rapidly and reached a larger volume compared with no vaporized PSNE. We also noted that lesions formed due to bubble-enhanced heating were more often prefocal due to backscattering from the bubble cloud. Our results show that these effects were more pronounced at higher PSNE concentrations. Finally, a comparison between the effects of pulsed and cw exposure on cavitation activity and lesion shape was investigated.

5:00

4pBB13. Observations of microbubble translation near vessel walls. Hong Chen, Wayne Kreider, Andrew A. Brayman, Michael R. Bailey, and Thomas J. Matula (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA)

To exploit the potential clinical applications of ultrasound contrast agents, it is important to understand the basic physics of bubble-vessel interactions. In this work, high-speed micrography was used to investigate

the dynamics of microbubbles inside vessels from *ex vivo* rat mesentery under the exposure of a single ultrasound pulse. The study included arterioles and venules. The ultrasound pulses were about 2 μs long with a center frequency of 1 MHz and peak negative pressures between 1 and 7 MPa. High-speed photographs reveal that interactions between microbubbles and vessels caused bubbles to move away from nearby vessel walls. The motion depended upon the standoff distance between the bubble and the nearest vessel wall, where a normalized standoff γ can be defined as the ratio of the standoff distance to the maximum bubble radius. For $\gamma < 1$, obvious bubble movement away from the nearby vessel wall was observed during collapse. At $\gamma = 0.3$, a bubble inside a venule was observed to translate over 16 μm during 1.5 μs, achieving velocities over 10 m/s. Bubble translation away from the vessel wall may be caused by the rebound of the vessel wall that was distended during bubble expansion. [Work supported by NIH Grants EB00350, AR053652, DK43881, and DK070618.]

5:15

4pBB14. Bubble-boundary interactions relevant to medical ultrasound. Wayne Kreider, Hong Chen, Michael R. Bailey, Andrew A. Brayman, and Thomas J. Matula (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, wkreider@u.washington.edu)

Because bubble-vessel interactions have been implicated in vascular damage associated with medical ultrasound, such interactions remain a topic of interest. Recent photographs have shown that bubbles inertially oscillating in blood vessels tend to cause a vessel invagination (displacement toward the lumen) that exceeds the corresponding distention. To gain an understanding of such interactions, a Bernoulli-type equation was derived for conditions in which flow near a bubble is confined by a nearby planar boundary. This formulation assumes incompressible, inviscid flow within a locally cylindrical geometry between the bubble and boundary. Using spherical, radial bubble dynamics inferred from the aforementioned photographs, pressures generated by the bubble at the boundary were calculated. The assumed sphericity was based on observations of bubble collapses and jetting behaviors. In addition, the form of the equation was compared to that for unconfined flow around a spherical bubble. The main conclusion is that tensile stresses corresponding to vessel invagination will be asymmetrically larger than compressive stresses associated with vessel distention as the amplitude of oscillation increases and the distance between the bubble and boundary decreases. As such, these tensile stresses may be an important mechanism for vascular bioeffects. [Work supported by NIH Grants EB00350, DK43881, and DK070618.]

5:30

4pBB15. PiggyBac transposon-based gene delivery with cationic ultrasound contrast agents into SCID and CD-1 mice. Pavlos Anastasiadis (Dept. of Mech. Eng., Univ. of Hawaii at Manoa, 2540 Dole St., Holmes Hall 302, Honolulu, HI 96822, pavlos@hawaii.edu), Terry O. Matsunaga (Univ. of Arizona, Tucson, AZ 85724), Stefan Moisyadi, and John S. Allen (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Efficient integration of a functional gene, along with suitable transcriptional regulatory elements, poses a fundamental difficulty for the application of successful gene therapy. Though inducible gene expression after somatic cell gene transfer has been achieved by employing viral vectors, issues associated with cargo capacity, host immune response, the risk of insertional mutagenesis, and the requirement for separate viruses are limiting factors. Several researchers have used transposon-based approaches for gene delivery, and the piggyBac transposon system has recently been shown to be effective in human cell lines and for the transgenesis in mice. The use of engineering helper-independent plasmids with the mouse codon-biased piggyBac transposon (mPB) gene in combination with cationic ultrasound contrast agents (UCAs) is investigated for a range of insonification parameters. The plasmid pmGENIE-3 deactivates the mPB gene after insertion of the transposon, eliminating potentially negative effects which may occur from the persistence of an active piggyBac gene post-transposition.

5:45

4pBB16. Quantification of endothelial cell transfection as a function of permeability with targeted ultrasound contrast agents. Pavlos Anastasiadis (pavlos@hawaii.edu) and John S. Allen (Dept. of Mech. Eng., Univ. of Hawaii at Manoa, 2540 Dole St., Holmes Hall 302, Honolulu, HI 96822)

Gene and drug delivery have shown to be enhanced by the localized destruction of ultrasound contrast agents (UCAs). Moreover, deoxyribonucleic acid (DNA) is rapidly degraded by serum DNases following its injection into the bloodstream, and several studies have shown that DNA may be protected from such DNases if it is conjugated onto the shell of UCAs. The optimal acoustic parameters for transfection and delivery are not well understood. Likewise, the real-time monitoring of the related cell membrane permeability changes has only been recently attempted for these applications. This study evaluates the endothelial cell-to-cell and cell-to-substrate gaps with the electric cell-substrate impedance sensing system (ECIS, Applied BioPhysics, Troy, NY). ECIS can detect in real-time the nanometer order changes of cell-to-cell and cell-to-substrate distances separately. Targeted UCAs were conjugated with plasmid vectors and subject to ultrasound insonification, while continuous measurements of the electrical resistance across the endothelial monolayers were conducted in real-time for the determination of membrane permeability changes. The achieved transfection rate is investigated over a range of acoustic parameters. [This work was supported by the National Institutes of Health Grants NIH 2 P20 RR016453-05A1 and NIH 2 G12 RR0030161-21.]

Session 4pEA

Engineering Acoustics and Signal Processing in Acoustics: Directional Microphone Arrays Signal Processing

James E. West, Chair
Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218

Chair's Introduction—1:00

Invited Papers

1:05

4pEA1. Optimal spherical array modal beamformer in the presence of array errors. Haohai Sun (Dept. of Electrons and Telecommunications, Norwegian Univ. of Sci. and Tech., 7491 Trondheim, Norway, haohai.sun@iet.ntnu.no), Shefeng Yan (Chinese Acad. of Sci., 100190 Beijing, China), and U. Peter Svensson (Norwegian Univ. of Sci. and Tech., 7491 Trondheim, Norway)

Spherical array modal beamforming has become an attractive technique, since three-dimensional beampatten synthesis is more flexible than other array geometries, and the modal beamforming can be performed using the elegant spherical harmonics framework. However, the performance of a modal beamformer in practical situations is known to degrade in the presence of array errors caused by non-perfect spatial sampling, sensor sensitivity and phase variations, and sensor self-noise. Although the white noise gain constraint has been widely used to control the robustness of modal beamformers, it is not clear how to properly choose the constrained parameter set based on the known range of errors. In this paper, a worst-case performance optimization approach is formulated for the spherical array modal beamformer to improve its robustness against the above mentioned errors occurring in practice; thus the optimum performance can be obtained based on the known maximum level of errors. The robust optimal modal beamforming problem is reformulated as a convex optimization within the spherical harmonics framework, which can be efficiently solved by using second order cone programming.

1:25

4pEA2. Analysis of the high-frequency extension for spherical eigenbeamforming microphone arrays. Jens Meyer and Gary W. Elko (mh acoustics LLC, 25A Summit Ave., Summit, NJ 07901)

The em32 Eigenmike® spherical microphone array is constructed with 32 omnidirectional pressure microphones mounted on the surface of a 8.4-cm rigid spherical baffle. Beamforming is accomplished by the spherical harmonic decomposition eigenbeamformer and modal-beamformer concept. Due to the array size, geometry, and average element spacing, this approach yields fully controllable beampattern control up to approximately 8 kHz. Above this frequency, spatial/modal aliasing begins to interfere with beampattern control. However, professional audio applications require an operating bandwidth to much higher frequencies. One solution is to use a single microphone capsule that is closest to the main axis of the beampattern direction at higher frequencies. It will be shown that there are some problems to this solution. As an alternative implementation it is shown that using a small subset of microphone capsules at high frequencies allows one to smoothly transition the beampattern from eigenbeamformer processing to higher frequencies.

1:45

4pEA3. Audio visual scene analysis using spherical arrays and cameras. Adam O'Donovan, Ramani Duraiswami, Dmitry Zotkin, and Nail Gumerov (PIRL Lab., Univ. of Maryland, 3366 AVW Bldg., College Park, MD 20742, adamod@gmail.com)

While audition and vision are used together by living beings to make sense of the world, the observation of the world using machines in applications such as surveillance and robotics has proceeded largely separately. We describe the use of spherical microphone arrays as "audio cameras" and spherical array of video cameras as a tool to perform multi-modal scene analysis that attempts to answer questions such as "Who?,", "What?," "Where?," and "Why?." Signal processing algorithms to identify the number of people and their identities and to isolate and dereverberate their speech using multi-modal processing will be described. The use of graphics processor based signal processing allows for real-time implementation of these algorithms. [Work supported by ONR.]

2:05

4pEA4. A comparison of sound localization methods in a room environment using a spherical microphone array. Colin Barnhill and James E. West (Dept. ECE, Johns Hopkins Univ., Baltimore, MD 21218, cb@jhu.edu)

Enclosures are one of the most challenging environments for source localization, especially for teleconferencing where the main goal is to enhance desired speech signals while suppressing noise. This can be accomplished using beamforming but the speech source locations must be known. Three passive methods will be compared for room environment localization: generalized cross-correlation (GCC), spherical harmonic-based multiple signal classification (MUSIC), and spatial power signal characteristics and gradients (SPSGs). SPSG is a new localization algorithm designed for room environments based on wideband third order beamforming and power

signal characteristics, differences, and gradients. MUSIC is a subspace method that relies on an eigen-decomposition of the beam correlation matrix and GCC is a time delay estimation algorithm based on microphone signal correlation. MUSIC and GCC are methods that perform well in their optimal environments but suffer in noisy and reverberant conditions. Results will be shown for multiple source situations with reverberation and noisy interferers.

Contributed Papers

2:25

4pEA5. Microphone array sensor node for distributed wireless acoustic beamforming. Alaa A. Abdeen and Laura E. Ray (Thayer School of Eng., Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755)

The presentation will illustrate the use of field programmable gate arrays (FPGAs) in microphone array digital signal processing for wireless acoustic beamforming in source localization. We have developed a prototype of a node composed of ten microphone channels interfaced with a Xilinx-based FPGA board. The methodology involves basic signal conditioning and delay-and-sum beamforming programmed by VHDL. The prototype results illustrate significant increase in the sensitivity to acoustic sources in the direction of a signal, while rejecting noise from other directions. The sensitivity is also enhanced in relationship to the number of channels. This node prototype is a building block for a wireless sensor network, capable of distributed beamforming at about 10-kHz sample frequencies. Each node is designed to maximize throughput so that broadband acoustic sources are accommodated. We are now performing parametric studies to optimize the node design and microphone array geometry to provide maximum bandwidth, dynamic range, and adequate special resolution, while preserving the ability to resolve time delays between the microphones.

2:40

4pEA6. MEMS (microelectromechanical systems) microphone array on a chip. Joshua S. Krause and Robert D. White (Mech. Eng., Tufts Univ., 200 College Ave., Medford, MA 02155, joshua.krause@tufts.edu)

The design, fabrication, and characterization of a surface micromachined, front-vented, 64 channel (8×8), capacitively sensed microphone array-on-a-chip are described. The element pitch is 1.3-mm center-to-center and is targeted at resolving the high-wavenumber components of the pressure spectra underneath the turbulent boundary layer. Care is taken to minimize package topology to reduce flow generated self-noise. The array was fabricated using the MEMSCAP PolyMUMPs® process, a three layer polysilicon surface micromachining process, with the addition of a Parylene-C coating in post-processing. An acoustic lumped element model, including mechanical components, environmental loading, fluid damping, and other acoustic elements, is detailed. Laboratory calibration indicates a sensitivity of 1 mV/Pa for each microphone over a 200–40 000 Hz band. A strong resonance occurs at 280 kHz in close accordance with modeled results. Spatial mapping of the array reveals minimal electrical and physical cross-talk between elements. Preliminary element-to-element sensitivity comparisons exhibit a standard deviation of 13%, 12%, and 25% at 250 Hz, 500 Hz, and 1 kHz. Phase matching showed a 9, 10, and 15 deg standard deviation difference at the same frequencies. A more in depth analysis of acoustic sensitivity is ongoing as well as element-to-element variability.

2:55

4pEA7. Ultrasonic calibration of MEMS (microelectromechanical systems) microphone arrays. Katherine E. Knisely, Victor Singh, Karl Grosh, David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109, kknisely@umich.edu), and Serdar H. Yonak (Toyota Motor Eng. and Manufacturing, North America, Inc., Ann Arbor, MI 48105)

Ultrasonic microphone arrays have applications in imaging, scale model testing, and sound source localization. For these applications MEMS ultrasonic microphones are emerging as a natural choice because of their small size, low cost, and robustness to manufacturing treatments and environmental conditions. In this talk an array is constructed using commercially available Knowles SiSonic[TM] capacitive MEMS microphones. The absolute sensitivity and phase calibrations of the array elements in the 1–60-kHz frequency range are obtained through the use of replacement calibration and time-frequency analysis. For these results, a modulation in sensitivity and phase measurements for frequency ranges over 15 kHz is observed under certain conditions. The origin of this modulation is discussed as well as various techniques for its removal. Relative element calibrations are crucial for array performance and are obtained through a variety of sound source configurations, which vary in size and location relative to the array.

3:10

4pEA8. Characterization of a high-frequency pressure-field calibration method. Dylan Alexander (Mech. and Aero. Eng., Univ. of FL, 231 MAE-A, Gainesville, FL 32611, fhdpa86@ufl.edu), Casey Barnard, Benjamin A. Griffin, and Mark Sheplak (Univ. of Florida, Gainesville, FL 32611)

A method for the direct measurement of the pressure sensitivity of a microelectromechanical system (MEMS) microphone is developed for frequencies from 10 to 100 kHz. The use of a simultaneous pressure field measurement reduces the errors in calculating the pressure sensitivity from free field measurements, and it allows calibration of non-reciprocal and non-capacitive transduction schemes. An ionophone is used as a calibration source in an anechoic chamber. An aluminum plate is mounted in the anechoic chamber and contains a linear array of pressure field condenser microphones. Treatment of the plate edge is used to minimize diffraction within the pressure field. The microphone data are used to characterize the directivity of the ionophone and localize sources of acoustic diffraction that could impact future MEMS microphone magnitude and phase calibrations.

3:25

4pEA9. Experimental study of synthetic aperture microphone array on rotor blade noise. Jaehyung Lee, Wook Rhee, and Jong-Soo Choi (Dept. of Aerosp. Eng., Chungnam Nat'l. Univ., Daejeon 305-764, Republic of Korea aerojhl@cnu.ac.kr)

Recent development of signal processing technologies has encouraged the aero-acousticians to implement the state-of-the-art sensors and data acquisition devices. They sometimes cost too much; however, especially for the academic arenas, the efforts of reducing the expense of integrating a system and increasing performance of phased array have initiated the idea of designing the synthetic aperture microphone array. The application of beamforming technique in localizing rotor blade noise and the use of synthetic aperture array in wind tunnel testing are presented. To apply the synthetic aperture beamforming on rotating objects, three steps should be considered. The first step is to restore the original acoustic waveform radiated from a rotating sound source. The second step is to synthesize independent array measurement data by correcting phase information of the received acoustic signals. Finally, the beamformer calculates the power on each point of grid in the space. The paper describes the details of deploying the synthetic aperture microphone array and the data reduction methods associated with acoustic measurement in wind tunnel. The comparisons of various patterns of synthetic aperture array are made. The beampower maps for rotor blade noise of various flight conditions are presented.

3:40

4pEA10. Comparison of single-frequency monopulse techniques that mimic the results of multiple-frequency, single-aperture interferometry. Daniel S. Brogan and Kent A. Chamberlin (Dept. of Elec. and Comput. Eng., Univ. of New Hampshire, 33 Academic Way, Durham, NH 03824-2619, daniel.brogan@unh.edu)

The capability to resolve targets from each other and from noise is desirable in all target detection and tracking systems. In a previous presentation [G. R. Cutter, Jr. and D. A. Demer, J. Acoust. Soc. Am. **126**, 2232 (2009)] it was shown that single-frequency, multiple-aperture interferometry could be used to resolve multiple targets that could not be resolved using single-frequency, single-aperture interferometry. This was accomplished by using the phase differences from several pairs of subarrays with different

subarray spacing to mimic the operation of a multiple-frequency, single-aperture interferometric system. The differential-phase technique is one type of phase-comparison monopulse used in target bearing estimation. The current research compares the results of single-frequency, multiple-aperture interferometry to those of other single-frequency phase-comparison and amplitude-comparison monopulse techniques by using multiple receive subarrays and/or varying the receive arrays' elements' weights.

3:55

4pEA11. Non-uniform array synthesis concept and theory. Dehua Huang (NUWC, Newport, RI 02841)

Uniform or symmetric array is a common practice for most of the sonar array design, where the same performance transducer elements are used for numerical simulation and manufacture engineering conveniences. However, non-uniform array designs show their advantages for special array applications. This paper will theoretically address the non-uniform array synthesis concept and theory. A few numerical examples will also be presented for element hybrid and cluster special arrays. [The work is supported by the U.S. Navy.

4:10

4pEA12. Comprehensive development of the theory of nonuniformly spaced arrays. Jenny Y. Au and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854)

In this work, we will present the work of comprehensive development of the theory of nonuniformly spaced arrays. It will be shown that the position may be determined in terms of the Fourier transform of the objective function. Special consideration will be given to the objective function of G. J. van der Maas in 1954. These results will be compared with that presented by Wallace in 1982.

THURSDAY AFTERNOON, 22 APRIL 2010 HARBORSIDE B, 2:00 TO 3:25 P.M.

Session 4pID

Interdisciplinary: Science and Human Rights Coalition: The Interface Between the Human Rights and Scientific Communities

Edward J. Walsh, Chair

Boys Town National Research Hospital, 555 N. 30th St., Omaha, NE 68132

Chair's Introduction—2:00

Invited Papers

2:05

4pID1. The Science and Human Rights Coalition of the AAAS (American Association for the Advancement of Science): The interface between the human rights and scientific communities. Jessica Wyndham (AAAS, Sci. and Human Rights Coalition, 1200 Washington Ave., NW, Washington, DC 20005)

In July 2005, AAAS held a 2-day conference of scientific associations and human rights organizations to explore ways in which the scientific community could become more directly engaged in human rights. Born of a conference recommendation, the AAAS Science and Human Rights Coalition was officially launched in January 2009. Forty-two scientific membership organizations have since joined the Coalition which aims to build bridges and opportunities for collaboration within the scientific community and between the scientific and the human rights communities. Introducing the Coalition, Ms. Wyndham will answer four questions: What does the Coalition aim to accomplish? How will it do this? Who can join? How does it operate? One year young, the growing number of Coalition members reflects the commitment of scientific organizations to addressing human rights. This, however, is just the beginning. Building on the achievements to-date and lessons learnt throughout the course of the first year, Ms. Wyndham will outline the Coalition's plans for making a real and meaningful contribution to the realization of human rights.

2:25

4pID2. Aspects of participation in the Science and Human Rights Coalition of the AAAS (American Association for the Advancement of Science): Working groups and internal outreach. Clinton W. Anderson (American Psychol. Assoc., 750 First St., NE, Washington, DC 20002)

The goal of Dr. Anderson's presentation is to inform ASA members about two practical aspects of participation in the AAAS Science and Human Rights Coalition: Participation in the working groups that are at the core of the Coalition's activities and outreach by coalition member representatives internally within their organizations. The speaker will briefly introduce the missions and modes of operation of the five working groups that are currently active within the Coalition and focus more extensively on the operations and objectives of the Service to the Scientific Community Working Group. Based on the action plan of the Service to the Scientific Community Working Group and his own experience as an organization member representative for the American Psychological Association, the speaker will to explore ways in which ASA members might work to increase awareness and appreciation of human rights issues and alternatives within the ASA; to increase capacity of the ASA to address human rights issues through a variety of means, including the application of acoustical tools and techniques; and to increase engagement of ASA in human rights activities.

2:45

4pID3. The Acoustical Society of America and the Human Rights Coalition: An introduction. Edward J. Walsh (Boys Town Natl. Res. Hospital, 555 North 30th St., Omaha, NE 68132)

Technological developments in recent years have significantly improved the capacity of human rights advocates to monitor and document abuses of authority, including violations of human rights at home and in remote isolated regions of the world. Recognizing the need for a platform from which scientists and scientific Societies might speak with a common voice to promote human rights protect the professional and personal interests of threatened colleagues, cooperate with human rights organizations, and press for the practical implementation of the right of all persons to the benefits of scientific progress and its applications, the AAAS Science and Human Rights Coalition was launched in 2008. Operating through the Panel on Public Policy, the ASA accepted an invitation to participate as an affiliate member of the Coalition, and the purpose of the Special Session is to introduce the membership at large to the principles and goals of the organization and to offer insight into the mechanism of operation and cooperation through which an effective interface between the scientific and human rights communities might be developed.

3:05

4pID4. The role of science in the human rights movement. Leonard Rubenstein (Ctr. for Public Health and Human Rights, Johns Hopkins School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205)

"Scientists have contributed valuably in making human rights a reality for all. Notable scientific contributions to human rights include the forensic exhumation of mass graves, the use of DNA evidence to identify victims of mass killings, the introduction of information management techniques to illuminate large-scale human rights violations, and, more recently, the use of satellite imagery to document the destruction of communities in remote locations." [Rubenstein and Younis 2008)]. As a human rights lawyer and a consumer of science, Mr. Rubenstein will discuss the contributions of science to human rights challenges, and the importance of the scientific community becoming a constituency for human rights. He will address head-on some of the greatest misconceptions about human rights work, including the charge that it is too "political" as well as steps needed to assure that scientific work undertaken with human rights organizations and is conducted consistent with scientific traditions of impartiality and independent inquiry. Finally, he will demonstrate how the AAAS Science and Human Rights Coalition can be used as a vehicle for the application of scientific skills, to reinforce the commitment, and to amplify the voices of scientists working toward the realization of human rights for all.

THURSDAY AFTERNOON, 22 APRIL 2010 ESSEX A/B/C, 3:15 TO 5:30 P.M.

Session 4pMU

Musical Acoustics: General Topics in Musical Acoustics

Thomas R. Moore, Chair
Rollins College, Dept. of Physics, Winter Park, FL 32789

Contributed Papers

3:15

4pMU1. Novel exploration of the vibration of the Caribbean steelpan in response to authentic excitation. Patrick O'Malley, Joseph F. Vignola, and John A. Judge (Dept. of Mech. Eng., The Catholic Univ. of America, 620 Michigan Ave., Washington, DC 20064)

A novel conformal scanning laser vibrometer was used to determine the vibrational modes of a C-lead tenor steelpan. The measurements provide a rich mechanical snapshot of the artistry involved in the marriage of metallurgy and music theory that goes into the creation of the instrument. The data represent the surface-normal motion of the entire face of the instrument in response to an impulsive excitation intended to mimic the strike of a mallet. A traditional scanning laser Doppler vibrometer is limited to measurement of planar geometries and therefore is of limited utility on the highly contoured surface of the steelpan. A description of the novel measurement system is presented, followed by the surface velocity data and a summary of response shapes and the frequencies at which those responses occur. The data show that individual note areas respond when adjacent or non-adjacent musically related notes are struck and clearly illustrate the complex vibration of the steelpan and the coupling between notes that produce the rich distinctive nature of the steelpan sound.

3:30

4pMU2. Vibrational characteristics of Balinese gamelan metallophones. Molly E. Jones (Dept. Phys., Univ. of Michigan, 450 Church St., Ann Arbor, MI 48109, jonesmo@umich.edu), Kent L. Gee, and Jeremy Grimshaw (Brigham Young Univ., Provo, UT 84602)

A study of the 16 metallophones from a Balinese gamelan has been conducted. Acoustical recordings of metallophone bars being struck were used to examine ratios of overtone frequencies to the fundamental. Results showed large variability in the number and ratios of overtones present. Scanning laser Doppler vibrometry measurements made on several bars also revealed great variability in mode shapes present. The distribution of prominent overtones and their modal shapes do not appear to match those of Western metallophones. Notably, the overall gamelan metallophone characteristics are quite dissimilar to the glockenspiel, which disagrees with previous studies. [This study was supported by the National Science Foundation through its Research Experiences for Undergraduates program.]

3:45

4pMU3. An introduction to the acoustics of playing the horn. David J Zartman (Graduate Program in Acoust., Penn State Univ., 405 EES Bldg., State College, PA, 16801)

In the art of music, a good performer can make a low-quality instrument sound passable, while a bad performer will not play well even on the best instrument in existence. Rather than merely focusing on the properties of the instrument, it is useful to provide the musician with some of the acoustical principles involved with playing the horn well. At the same time, this provides a medium to enable the acoustician to understand some of the intricacies a musician must balance well to create a pleasing sound.

4:00

4pMU4. Novel computer aided design of labial flue pipes. Brian Moss (CSRC/OFSRC, Univ. of Limerick, Limerick, Ireland), Elfed Lewis, Gabriel Leen, Kort Bremer, and Andrew Niven (Univ. of Limerick, Limerick, Ireland)

A labial flue pipe is a well known tone generator, which is familiar and easily recognizable as the organ pipes seen in many concert halls and churches. However, the design and understanding of the sounding mechanism of such pipes are fraught with difficulty. Traditionally labial pipes are constructed from age-old lookup tables that are closely guarded intellectual secrets. This paper discusses a novel computer program that facilitates the design and construction of such labial flue pipes. The computer program allows almost all aspects of the labial flue pipe design to be varied, the resultant frequency is generated, and in addition the Ising efficiency number is provided. Furthermore, a discussion is included related to the fact that even though an Ising number greater than 3 indicates that a pipe is overblown, the fundamental tone is still predominant. A comparison will also be made between a CFD simulation of the labial flue pipe jet mechanism and smoke trail plots typical of such analysis.

4:15

4pMU5. Acoustical studies on conch shells. M.G. Prasad and B. Rajavel (Dept. of Mech. Eng., Noise and Vib. Control Lab., Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad@stevens.edu)

Sound from blowing a conch shell is known for its tonal quality. Its use in many cultures is known. In Hindu culture, conch shell is blown on several auspicious occasions such as festivals, worship, and rituals. A conch shell can be modeled as a horn in its acoustical behavior. This work presents an acoustical study in terms of the spectral characteristics of the shell. The study also presents the influence of parameters such as placing hand in the mouth of the conch shell and the size of conch shell on the spectral characteristics.

4:30

4pMU6. On the role of acoustics in the Vedic Hindu tradition and philosophy. M. G. Prasad and B. Rajavel (Dept. of Mech. Eng., Noise and Vib. Control Lab., Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad @stevens.edu)

Acoustics has received a very high importance in the Vedic Hindu tradition and philosophy of ancient India. The Vedic literature which refers to both the Vedas and the subsequent literature based on Vedas emphasize that the roles of sound both as source signal and as hearing are important. This work presents the various ways such as chants, Sanskrit language, vocal, and instrumental music in which acoustics has played a major role.

4:45

4pMU7. The shape of musical sound: Real-time visualizations of expressiveness in music performance. Gang Ren (Dept. of Elec. and Comput. Eng., Edmund A. Hajim School of Eng. and Appl. Sci., Univ. of Rochester, Rochester, NY 14627), Justin Lundberg, Mark F. Bocko, and Dave Headlam (Univ. of Rochester, Rochester, NY)

Despite the complex multi-dimensional nature of musical expression in the final analysis, musical expression is conveyed by sound. Therefore the expressiveness of music must be present in the sound and therefore should be observable as fundamental and emergent features of the sonic signal. To gain insight into this feature space, a real-time visualization tool has been developed. The fundamental physical features—pitch, dynamic level, and timbre (as represented by spectral energy distribution)—are extracted from the audio signal and displayed versus time in a real-time animation. Emergent properties of the sound, such as musical attacks and releases, the dynamic shaping of musical lines, timing of note placements, and the subtle modulation of the tone, loudness, and timbre can be inferred from the fundamental feature set and presented to the user visually. This visualization tool provides a stimulating music performance-learning environment to help musicians achieve their artistic goals more effectively. Such visualizations and interactions with musical sound also will promote the semantic understanding of an expressive musical language.

5:00

4pMU8. Robust audio watermarking scheme based on spectral modeling synthesis. Pranab Dhar (School of Comput. Eng. & Information Technol., Univ. of Ulsan, Korea, pranab_cse@yahoo.com), Cheol Hong Kim (Chonnam Natl. Univ., Chonnam, Korea), and Jongmyon Kim (Univ. of Ulsan, Korea)

This paper proposes a new watermarking scheme based on spectral modeling synthesis for copyright protection of digital contents. In our proposed scheme, a short time Fourier transform (STFT) is applied to the original signal. Prominent spectral peaks of each frame are then detected and represented by sinusoidal parameters. The residual component is computed by removing all the prominent peaks from the spectrum, transforming the spectrum back to the time domain using inverse fast Fourier transform, and overlap-adding the frames in time. In addition, a peak tracking unit keeps the sinusoidal parameters and forms peak trajectories. Watermarks are then embedded into the most prominent peak of each frame in the peak trajectories. Watermarked signal is computed by the sinusoidal synthesis process and then adding the sinusoids with the residual signal. Watermarks are detected by using the inverse operation of watermark embedding process. Simulation results indicate that the imperceptible watermarks embedded into four different kinds of music sounds using the proposed scheme are robust against several attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression, showing similarity values ranging from 0.8 to 0.9. [Work supported by KOSEF, R01-2008-000-20493-0.]

5:15

4pMU9. Effects of tempo and other musical features on stress responses to college students. Wei-Chun Wang (Dept. of Digital Lit. and Arts, St. John's Univ., 499, Sec. 4, Tam King Rd., Tamsui, Taipei, Taiwan, vgnwang @hotmail.com)

It is believed that music has the power to soften emotions and alleviate pains. This study is aimed to explore the effects of tempo and other musical elements on stress-associated responses of college students when they listen to music. The effects of musical experience, preference, and awareness of music content on stress ratings are also investigated. College students were asked to listen to eight recorded musical excerpts and fill in a questionnaire to rate their relaxation and anxiety levels in a Likert five-point scale and to specify the musical elements. With a variety of tempo being considered, the musical excerpts were chosen from different types of music, including western classical music, Chinese pop music, Chinese traditional music, rock music, jazz blue, and new age music. Preliminary results showed that there was a correlation between relaxation and tempo, and the degree of personal music preference promoted stress reduction. It suggests that music used to enhance relaxation should match a person's musical taste.

Session 4pPA

Physical Acoustics: Nonlinear Acoustics

Martin D. Verweij, Cochair
Delft Univ. of Technology, Delft, The Netherlands

Robin O. Cleveland, Cochair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Contributed Papers

1:00

4pPA1. Miniature thermoacoustic engine: Experiments and modeling. Konstantin I. Matveev and Sungmin Jung (MME School, WSU, Pullman, WA 99164-2920, matveev@wsu.edu)

A miniature standing-wave thermoacoustic engine was constructed and tested with four quarter-wavelength resonators having lengths in the range of 57–124 mm. 80-ppi reticulated vitreous carbon (RVC) foam was applied as a stack, and an atmospheric air served as a working fluid. The critical temperature difference between the stack ends, corresponding to the sound onset, was recorded in the range of 200–300 °C. A thermoacoustic model was developed to predict the critical temperature difference across the stack. The model allows for user-defined thermoacoustic functions for the stack. Results obtained with the capillary-based theory for tortuous media and thermoacoustic functions previously measured for RVC demonstrate a trend qualitatively consistent with our data, while the quantitative agreement is lacking. Modeling results obtained with thermoacoustic functions for a parallel-plate stack can approximate experimental data, if the appropriate plate spacing is selected. [Work supported by the NSF Grant No. 0853171.]

1:15

4pPA2. Characterization of the heat flux through the heat exchangers of a thermoacoustic cooler. Gaëlle Poignand, Philippe Blanc-Benon, Emmanuel Jondeau (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France, gaelle.poignand@ec-lyon.fr), Etienne Gaviot, Lionel Camberlein, Guillaume Penelet, and Pierrick Lotton (l'Universit du Maine, 72085 Le Mans, France)

A small-scale standing wave thermoacoustic cooler with a couple of stack and heat exchangers is studied. In addition to classical instrumentation in such a device, thermal heat flux sensors specifically developed using MEMS technology equipped the heat exchangers of this refrigerator. These sensors give the temporal evolution of the heat fluxes through the hot and the cold heat exchangers. Hence, they provide a better understanding of heat transfer between the stack and the heat exchangers. In this work, the temporal evolution of the temperature along the stack and of the heat flux through the heat exchanger is measured versus the acoustic pressure. The results show that for high-pressure levels, the heat flux extracted at the cold exchanger rapidly increases until a maximum value and then stabilizes at a lower value. The origin of this limitation may come from the formation of vortices behind the stack highlighted in the work of Berson *et al.* [Heat Mass Transfer **44**, 1015–1023 (2008)]. This work is supported by ANR (project MicroThermAc NT051_42101).

1:30

4pPA3. Inertial cavitation threshold dependence on high static pressures. Kenneth, B. Bader, Joel Mobley, Jason Raymond (Natl. Ctr. for Physical Acoust., Coliseum Dr, Univ., MS 38677, kbader@olemiss.edu), and D. Felipe Gaitan (Impulse Devices, Inc., Grass Valley, CA 95945)

The creation of a bubble in a liquid is energetically favorable when the liquid is subject to a net/an overall tension less than of its saturated vapor pressure. This gain is offset by the creation of a liquid-vapor/gas interface, and only bubbles larger than some critical size will spontaneously grow. The magnitude of the acoustic pressure required to produce a cavity of the critical size is termed the cavitation threshold of the liquid. The dependence of the cavitation threshold on hydrostatic pressure has previously been reported up to 130 bars in terms of electrical power applied to the acoustic driving transducer. These measurements used a standing wave set-up in a stainless steel spherical resonator (24.1 cm outer diameter, 1.9 cm thick) with a $Q >$ 10 000 when fluid loaded. This work will extend into higher pressure regimes and will be used to extrapolate the cavitation threshold at 1 kbar.

1:45

4pPA4. Determination of bubble size distributions in an ultrasonic cavitation field. Stéphane Labouret and Jacques Frohly (Département d'Opto-Acousto-Electronique, Institut d'Electronique, Micro-électronique et Nanotechnologies (IEMN), Université de Valenciennes et du Hainaut-Cambresis, le mont Houy, F-59 313 Valenciennes, France, labouret@laposte.net)

An electromagnetic method is proposed for determining bubble size distribution in the range 2–12 m in ultrasonic cavitation fields. The method is used to investigate the effect of the continuous and pulsed irradiations and of the presence of increasing concentration of a surfactant (sodium dodecyl sulfate) on the bubble size distribution in a standing wave ultrasound field. The electromagnetic sizing technique shows considerable promise as a means of tracking changes in the bubble size distribution during insonation and determining the role of such changes in transitioning from a field that contains a few discrete bubbles to one that exhibits widespread cavitation activity.

2:00

4pPA5. Pressure, temperature, and dissolved gas dependence of dielectric breakdown in water. Jonathan Sukovich and R. Glynn Holt (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

It has been shown experimentally that the optical breakdown strength of water is a pressure dependent quantity, growing with increasing pressure. The dependence of the breakdown strength on temperature and dissolved gas concentration over a larger range of pressures will be observed. Using a custom fabricated pressure vessel and high-power Nd:YAG laser, breakdown events will be generated and observed over a range of pressures from 0 to 25 kpsi. Observations of breakdown events will be made using a high-speed photodetector located behind the pressure vessel's optical windows. Dissolved gas concentration will be controlled and varied using a custom water preparation system over a range from water's vapor pressure (~20 torr) to atmospheric pressure. Temperature will be monitored using a thermocouple attached to the pressure vessel and the temperature dependence will be measured over a range from 20 to 35 °C. A comparison between current single detector methods and previous imaging methods of using breakdown to determine absolute pressure will then be made. [Work supported by Impulse Devices, Inc.]

2:15

4pPA6. A study on nonlinear ultrasonic waves in bubbly media: One and two-dimensional numerical experiments. Vanhille Christian (Escet, Univ. Rey Juan Carlos, Tulipan s/n, 28933 Mostoles Madrid, Spain, christian.vanhille@urjc.es) and Campos-Pozuelo Cleofé (Instituto de Acustica, CSIC, 28006 Madrid, Spain, ccampos@ia.cetef.csic.es)

Nonlinear ultrasonic waves are studied in bubbly media, typically liquids with gas bubbles. The acoustic behavior is modeled by a differential system which couples the acoustic pressure and the nonlinear dynamic of the bubbles. The differential equations are solved by means of a numerical model that gives the acoustic pressure and bubble volume variation as time traces (SNOW-BL code) in two configurations: one and two dimensions. Results are shown for one-dimensional progressive and standing waves as well as for two-dimensional progressive waves, with or without focus. Experiments are carried out with bubbles evenly distributed or forming clouds, for example, in the pre-focal zone. [This work is part of the research project DPI2008-01429.]

2:30—2:45 Break

2:45

4pPA7. Temporal evolution of laser-nucleated bubble clouds in an acoustic resonator. Phillip A. Anderson and R. Glynn Holt (Dept. Mech-Eng, Boston Univ., 110 Cummington St., Rm. 101, Boston, MA 02215)

Using high-speed digital imaging, the evolution of laser-nucleated bubble clouds over multiple acoustic cycles is observed. Five 7×7 custom phase gratings are employed to produce 245 beams, which are then focused into the center of a spherical resonator capable of high static and acoustic pressure. The dependence of the evolution of the bubble cloud on several system parameters is measured. Notably the resulting cloud(s) depend strongly on the phase of the laser firing, the laser energy per pulse, and the acoustic pressure. Number of bubbles, radius of individual bubbles, and effective radius of clouds will be reported as functions of time. The interplay between cloud dynamics and shock waves will also be discussed. [Work supported by Impulse Devices, Inc.]

3:00

4pPA8. Theoretical analyis of Schwinger's conjecture on sonoluminescence in relation to the boosting of sonluminescent transduction efficiency. Harvey C. Woodsum (Sonetech Corp., 10 Commerce Park North, Unit 1, Bedford, NH 03110)

In a previous paper [Spring ASA, Salt Lake (2007)], we have considered the potential viability of Schwinger's conjecture that sonoluminescent light radiation, which results from acoustically developed cavitation, derives from zero-point energy of the vacuum. In our prior paper, we considered the nonlinear interaction of zero-point electromagnetic modes in a cavitating bubble, through the Euler–Heisenberg theory for the scattering of light by light. Good agreement was found between theory and experiment with regard to both the efficiency and spectrum of the sonoluminescent radiation generated using a single acoustic frequency. We now consider a comparison of this theory with other experiments [Phys. Rev. Lett. **81**, 1961–1964 (1998)] which have demonstrated that cavitation bubbles created with acoustic waves having both first and second harmonics present, and having particular amplitude and phase arrangements, can result in sonoluminescent transduction efficiency increases of up to 300%. At this point in time, the predictive power of the current theory appears promising.

3:15

4pPA9. Nonlinear poroacoustic flow in rigid porous media. Pedro Jordan and Jim Fulford (U. S. Naval Res. Lab., Stennis Space Ctr., MS 39529)

An acoustic acceleration wave is defined as a propagating singular surface (i.e., wavefront) across which the first derivatives of the velocity, pressure, or density exhibit jumps. In this talk, the temporal evolution of the amplitude and the propagation speed of such waves are investigated in the context of nonlinear acoustic propagation in rigid porous media. By considering the exact conservation/balance equations, it is shown that there exists a critical value, the constant $\alpha^*(>0)$, of the initial jump amplitude such that the acceleration wave magnitude either goes to zero, as $t\to\infty$, or blows up, in finite time, depending on whether the initial jump amplitude is less than or greater than α^*. In addition, stability is addressed; a connection to traveling wave phenomena is noted, for which an exact traveling wave solution is obtained; and a comparison with the linearized case, i.e., the well-known damped wave equation, is also presented. Finally, the numerical solution of an idealized, nonlinear initial-boundary value problem involving sinusoidal signaling in a fluid-saturated porous slab is used to illustrate the finite-time transition from acceleration to shock wave, which occurs when the initial jump amplitude exceeds α^*. [Work supported by ONR/NRL funding.]

3:30

4pPA10. Different regimes of nonlinear pulse propagation in porous medium. Diego Turo and Olga Umnova (Acoust. Res. Ctr., Univ. of Salford, Salford, Greater Manchester M5 4WT, United Kingdom)

High-amplitude pulse propagation in rigid porous media has been investigated numerically and experimentally. At high-sound levels, strong interactions between different spectral components of the pulse make any frequency domain models difficult to use, therefore a time domain approach has been applied for the present research. The effect of Forchheimer nonlinearity (i.e., flow resistivity growth with particle velocity amplitude) in porous media is well studied. However, much less is known about the influence of memory effects on high-amplitude pulse propagation. The aim of this work is to study their relevance for high-amplitude pulses of different durations. A numerical finite difference time domain scheme has been developed which accounts for second order convection nonlinearity, Forchheimer correction, and memory effect simultaneously. It is shown that Forchheimer nonlinearity dominates for longer duration high-amplitude pulses, while convection terms and memory effect contribution become noticeable for shorter pulses of moderate amplitude. The numerical results are confirmed in a series of experiments with different duration pulses in the amplitude range from 120 Pa to 40 kPa.

3:45

4pPA11. Directivity and frequency control of an intense underwater laser acoustic source for navy applications. Melissa Hornstein, Theodore Jones, Antonio Ting, and Michael Nicholas (Acoust. Div., U.S. Naval Res. Lab., Washington, DC 20375)

We develop an intense laser acoustic source, in which a tailored laser pulse can compress underwater at a predetermined remote location. Optical compression results in laser-induced breakdown (LIB), localized heating, and acoustic shock generation. Recent experiments include near-field acoustic source characterization using lens-focused 400-, 800-, 532-, and 1064-nm pulses of Ti:sapphire and Nd:YAG lasers. Sound pressure levels over 215-dB were achieved using a compact laser. We have demonstrated control of the shape of the LIB plasma volume, and thereby control of the acoustic frequency spectrum and acoustic source directivity. By superposition of volume elements within the LIB, the acoustic pulse duration in a given propagation direction is determined by the parallel dimension of the LIB, divided by the parallel acoustic transit time across the LIB. Thus, the shape of the LIB strongly affects the acoustic pulse duration and directivity, and aspherical LIB volumes result in strongly anisotropic acoustic sources. In our experiments, the LIB volume shape was varied by laser pulse length, energy, optical bandwidth, and focusing angle. We also studied acoustic propagation in a 30 000-gal bubbly salt water tank. [This work is supported by the U.S. Office of Naval Research.]

4:00

4pPA12. Investigation of acoustic streaming in the cochlea. Katherine Aho, Elaine Vejar, and Charles Thompson (Dept. Elec. Eng., CACT, FA 203, 1 University Ave, Lowell, MA 01854)

In this work we will analyze acoustic streaming in the cochlea. A model will be developed to examine the steady flow induced by the harmonic excitation of a fluid in a rigid two-dimensional waveguide. The strength of the flow will depend on the time-phase between the axial velocity and the

boundary velocity. It was found in previous work that the amplitude of the flow is directly related to the streaming Reynolds number. These results will be used to compare with Bekesy's observations.

4:15

4pPA13. Acoustic performance predictions via a general theory for the scattering of sound by sound with experimental data from an operational parametric sonar system. Harvey C. Woodsum (Sonetech Corp., 10 Commerce Park North, Unit 1, Bedford, NH 03110)

In results previously described [J. Acoust. Soc. Am. 95, 2PA14 (1994)], a general theory for the scattering of sound by sound has been developed as an exact solution to the Lighthill–Westervelt equation of nonlinear acoustics. Most recently, a computer model based on this theory has been developed in order to support design analysis and performance prediction for parametric array echo-ranging systems and has been applied to a currently deployed operational parametric sonar. The present theory and model have the potential to support a wide range of analyses, including development of beam patterns, source levels, transient waveform effects, as well as the analysis of scattering of sound by sound from intersecting beams as well as the prediction of performance of parametric sources having multiple simultaneous transmit beams. Both the theoretical basis for the model and agreement with experimental data are presented.

4:30

4pPA14. Imaging multiple masked nonlinear scatterers applying a combination of time reversal principles and the selective source reduction method. Siegfried Vanaverbeke, Koen Van Den Abeele (Wave Propagation and Signal Processing, K.U. Leuven Campus Kortrijk, E. Sabbelaan 53, B-8500 Kortrijk, Belgium, koen.vandenabeele@kuleuven-kortrijk.be), Brian E. Anderson (Brigham Young Univ., Provo, UT 84602), and Marco Scalerandi (Politecnico di Torino, 10129 Torino, Italy)

Inherent limitations of the traditional time reversal process in the case of multiple sources or scatterers make it impossible to distinguish them individually. The selective source reduction (SSR) method employs a subtraction technique to selectively suppress in amplitude (and ideally eliminate) a time reversed focal signal that is masking another focus. In previous work, Scalerandi *et al.* and Anderson *et al.* successfully applied the SSR method to identify masked primary sources in a fully linear medium. Here, we extend the capabilities of the SSR method to deal with (i) secondary sources (for example, scattering caused by embedded defects) and (ii) nonlinear scattering generated during the ultrasonic wave propagation as well. We call this new method selective source reduction based on nonlinear time reversed acoustics (SSR-NLTRA). In the extended approach, the contribution of all primary sources is first eliminated by means of the scaling subtraction method. Subsequently, the SSR-TRA method is applied to the remaining nonlinear content of the signals. We show by means of two dimensional wave propagation simulations that the new method can be applied iteratively to successfully image multiple masked nonlinear defects.

THURSDAY AFTERNOON, 22 APRIL 2010 GRAND BALLROOM VII/VIII/IX, 2:00 TO 6:00 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Physiology, Models, and Auditory Processing (Poster Session)

Monita Chatterjee, Chair
Univ. of Maryland, Dept. of Hearing and Speech Science, College Park, MD 20742

Contributed Papers

All posters will be on display from 2:00 p.m. to 6:00 p.m. To allow contributors an opportunity to see other posters, contributors or odd-numbered papers will be at their posters from 2:00 p.m. to 4:00 p.m. and contributors or even-numbered papers will be at their posters from 4:00 p.m. to 6:00 p.m.

4pPP1. Vacuum ear plug. Martin L. Lenhardt (Dept. of Biomedical Eng., Virginia Commonwealth Univ., P.O. Box 980168 MCV, Richmond, VA 23298-0168, lenhardt@vcu.edu)

A passive inexpensive hearing protection device (HPD) will be described that provides hearing protection, which is simple to use and comfortable over hours of use. Protection from impact sounds will be provided by creating a negative pressure between the plug tip and the eardrum (−100-mm H2O or −1 kPa). The negative pressure will be created by squeezing the peripheral plug end, evacuating a fixed amount of air from the canal. The aperture diameter is 0.010 in. In effect the pump plug will have a similar acoustical action on the eardrum as a normal stapedius muscle contraction, but without its limitations (too slow and fatigue). An external flexible bladder forms the end which is held in the fingers to be inserted into the ear canal. The vacuum mechanism consists of bulb on a polypropylene frame embedded in a closed cell urethane flanged housing. With a negative pressure of −2 kPa the transmission loss is 20–25 dB. It is estimated that a negative pressure of −1 kPa would result in a transmission loss of about 15 dB. The vacuum effect is maximal in the low frequencies which will also attenuate body conducted sounds by reducing ossicular inertial.

4pPP2. Development and validation of a computational model of bone-conducted sound transmission for improved hearing protection design. Odile H. Clavier (Creare Inc., 16 Great Hollow Rd., Hanover, NH 03755, ohc@creare.com), Margaret Wismer (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801), Jed Wilbur, Anthony Dietz (Creare Inc., Hanover, NH 03755), and William O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Bone-conducted sound is the limiting factor in current hearing protection for very high-noise environments such as those encountered by maintainers of military aircraft. The University of Illinois has developed a three-dimensional acoustic wave propagation model for the computation of sound transmission into, around, and through fluid and solid-shell bodies. Creare has implemented experimental techniques to validate the computational

results. Three cases are presented. The first is a fluid sphere, for which experimental, computational, and analytical results were obtained. The second is a solid-shell, fluid-filled sphere, for which experimental and computational results were obtained to determine the effect of the solid shell. The third is a representative human head with and without passive hearing protection. For this case, the computational model and an experimental head simulator were both developed from a computed tomography scan of a living human head. The head simulator was built around a rapid-prototype instrumented skull using silicon organs and simulated tissue. Experimental, computational, and analytical results were all in good agreement for the fluid sphere test case. While general agreement was obtained for the other two cases, specific discrepancies in the results are outlined and limitations of the models are discussed.

4pPP3. Application of matched asymptotic expansions to the analysis of compressional bone conduction. David Chhan and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, david_chhan@student.uml.edu)

In this work, we will investigate the compressional mode of bone conduction and its effect on the displacement of the cochlear partition. We will use the method of matched asymptotic expansions to obtain the velocity of the partition. In this mode of bone conduction, the cochlea is compressed alternatively due to an applied vibratory force. Velocity difference between the oval and round windows is shown to be critical in determining the amplitude of the partition motion. From this physical process, we will analytically derive equations that will be used to solve for the velocity of the cochlear partition due to the compressional velocity on the cochlear shell.

4pPP4. Exploring the traveling wave dispersion in the cochlea. Sripriya Ramamoorthy, Ding-Jun Zha, and Alfred Nuttall, L (Dept. of Otolaryngol ./Head and Neck Surgery, Oregon Hearing Res. Ctr., Oregon Health and Sci. Univ., NRC 04, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, ramamoor@ohsu.edu)

In the measurements of the basilar membrane velocity as well as the auditory nerve responses to acoustic stimulation, it has been observed that low frequencies arrive at the measurement location (or the tonotopically located auditory nerve fiber) earlier and high frequencies arrive later, with significant delay noticed in the characteristic frequency component. In the derived impulse responses as well as click responses, this phenomenon manifests as the instantaneous frequency glide from low to high frequencies. The origin of these frequency glides has not yet been satisfactorily explained. In this paper, a simple elucidation along with experimental validation using measurements made on guinea pigs is presented for the plausible origin of the frequency glides observed in the cochlea.

4pPP5. Observing chartacteristic frequency shifts using chirp-evoked otoacoustic emissions. Luke Shaheen, Michael Epstein, and Ikaro Silva (Dept. Speech-Lang. Path. and Audiol., Auditory Modeling and Processing Lab. (AMPLab), Ctr. for Comm. and Digitial Signal Processing (Dept. Elec. and Comp. Eng.), Northeastern Univ., 360 Huntington Ave., Boston, MA 02115, m.epstein@neu.edu)

Otoacoustic emissions (OAEs) are often used to assess the integrity of specific regions of the cochlea. There is, however, evidence that the frequency that causes maximal response for a particular cochlear region shifts toward lower frequencies at high-stimulus levels. Therefore, OAE-based assessments of cochlear activity over a wide range of levels may need to take into account these shifts. The present study employed two experiments in order to confirm and characterize the extent to which this shift in cochlear excitation as a function of level can be observed in OAEs. In the first experiment, chirp-evoked otoacoustic emissions (ChOAEs) were used to track changes in the frequency spectra of each listener. The results showed a shift toward low frequencies in the spectral peak of ChOAE response in a majority of listeners (−9.1% average shift). This shift was then confirmed using a single-listener measurement of tone-burst otoacoustic emission input/output functions measured at three different frequencies. These results support the contention that it may be necessary to vary stimulus frequency to take the cochlear excitation shift into account when making assessments of the cochlear activity at a particular location across a wide range of levels.

4pPP6. Growth of otoacoustic emission suppression as a function of frequency and level in humans. Michael P. Gorga, Stephen T. Neely, Judy, G. Kopun, and Hongyang Tan (555 N. 30th St., Omaha, NE 68131)

DPOAE suppression was measured for 11 suppressor frequencies (f3) surrounding each of eight f2 frequencies (0.5–8 kHz) and six L2 levels (10–60 dB SL). A total of 63 normal-hearing subjects participated, with data from 20 subjects at each f2, L2 combination. Measurement-based stopping rules were used, such that averaging continued until the noise −25 dB SPL, the SNR 25 dB, or 210 s of artifact-free averaging had been completed. These stopping rules resulted in reliable measurements, even for conditions in which the SNR typically is poor (low f2 frequencies and low L2 levels). Suppression growth as a function of f3 was similar across f2 frequencies and L2 levels. Specifically, low-frequency suppressors relative to f2 had higher suppression thresholds but more rapid growth of suppression, compared to suppressor frequencies close to f2. Suppression growth for f3 frequencies above f2 was slow, even in comparison to suppression growth when f3 ¡Ö f2. This overall pattern was evident for all f2 frequencies and for a wide range of L2 levels. These findings suggest that suppression growth is similar across a wide frequency range in humans. [Work supported by the NIDCD R01 2251 and P30 4662].

4pPP7. Otoacoustic emission suppression tuning curves in humans. Michael, P. Gorga, Stephen, T. Neely, Judy, G. Kopun, and Hongyang Tan (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

DPOAE suppression data as a function of suppressor level (L3) for f2 frequencies from 0.5 to 8 kHz and L2 levels from 10 to 60 dB SL were used to construct suppression tuning curves (STCs). DPOAE levels in the presence of suppressors were converted into decrement versus suppressor level (L3) functions, and the L3 resulting in 3-dB decrements was obtained by transformed linear regression. These L3 levels were plotted as a function of f3 to construct STCs. STCs when f3 is plotted on an octave scale were similar in shape across f2 frequency. These STCs were analyzed to provide estimates of gain (tip-to-tail differences) and tuning (QERB). Both gain and QERB decreased as L2 increased, regardless of f2, but the increase with f2 was not monotonic. A roughly linear relation was observed between gain and QERB at each frequency, such that gain increased by 4–13 dB (mean xa1;xd6; 8 dB) for every unit increase in QERB, although the pattern varied with frequency. These findings suggest consistent nonlinear processing across a wide frequency range in humans, although the nonlinear operation range is frequency dependent. [Work supported by the NIDCD R01 2251 and P30 4662].

4pPP8. Influence of hair bundle configuration on biomimetic hair sensor sensitivity. Shuangqin Liu and Robert White (Mech. Eng., Tufts Univ., Medford, MA 02155, shuangqin.lin@tufts.edu)

Hair cells are the sensory receptors of the auditory system. When the stereocilia (hair buddles) sitting on top of the hair cells are stimulated, a current flows through the hair cell transducing mechanical stimuli into electrical response. For the outer hair cell (OHC), the hair bundles are stimulated by displacement of the tectorial membrane relative to the reticular lamina. For the inner hair cell (IHC), the hair bundles are stimulated by force due to motion of the fluid around the hair bundle. In addition, OHC and IHC hair bundles are shaped differently. In this paper, biomimetic micromachined hair sensors are designed, fabricated, and tested to investigate how the configurations of the hair bundles affect the sensitivity of the hair sensor. These hair sensors use a capacitive scheme and have different hair configurations including the W and U shape representing OHC and IHC hair bundles. Computational results achieved to date indicate that sensitivity can be strongly affected by the placement of the engineered hair bundles. Microfabrication and experimental work are ongoing.

4pPP9. Spatiotemporal coding of signals in the auditory periphery. Ramdas Kumaresan, Vijay Peddinti (Dept. of Elec. and Comput. Eng., Univ. of Rhode Island, Kelley Hall, Kingston, RI 02881), and Peter Cariani (Harvard Med. School, Newton, MA 02460)

Signal representation in the cochlea is often thought to involve either rate-place profiles or purely temporal, interspike interval codes. Spatiotemporal coding strategies based on phase-locking, cochlear delays, and coincidence detectors have also been proposed [Loeb *et al.*, Biol. Cybern. (1983); K. & Shamma, J. Acoust. Soc. Am. **107** (2000); and Carney *et al.*, Acoustica **88**, 334–337 (2002)]. In this view, spatiotemporal patterns of spikes locked to relative phases of the traveling wave at specific cochlear places at a given time can convey information about a tone. We propose a general mathematical basis for using such spatial phase/amplitude patterns along the frequency axis to represent an arbitrary (approximately) time and bandwidth-limited signal. We posit that the spatial pattern of phases and amplitudes corresponds to locations at which (real and/or imaginary parts of) the Fourier transform of the signal crosses certain levels (e.g., zero level). Given these locations, we show that we can accurately reconstruct the original signal by solving a simple eigenvalue problem. Using this approach, we propose an analysis/synthesis algorithm to represent speech-like signals. We conjecture that a generalized representation of the forms of signals can be inferred from spatial, cross-CF patterns of phase relations present in the auditory nerve. [Work supported by AFSOR FA9550-09-1-0119.]

4pPP10. Perception and discrimination of synchronous and asynchronous tones. Magdalena Wojtczak, Andrew J. Oxenham, and Anna C. Baird (Dept. of Psych., Univ. of Minnesota, Elliott Hall N218, 75 East River Rd., Minneapolis, MN 55455)

Studies of auditory-brainstem responses show that the response to a click is less synchronized than the response to an upward-chirp designed to compensate for the dispersion of basilar-membrane traveling-wave [Dau *et al.*, J. Acoust. Soc. Am. **107**, 1530–1540 (2000)]. However, upward-chirps are perceived as less compact than the clicks despite producing a more synchronized response along the basilar-membrane, suggesting the existence of a mechanism compensating for different traveling-wave delays [Uppenkamp *et al.*, Hear. Res. **158**, 71–83 (2001)]. This study evaluated synchrony and asynchrony perception and discrimination using tonal stimuli that excited remote places along the basilar membrane. Potential within-channel cues were masked using noise bands that were geometrically centered between the test tones. Level effects were investigated using two levels of the test tones, 20-dB SL and 85-dB sound pressure level (SPL). In addition, perception and discrimination of synchrony and asynchrony were measured for tones presented in noise bands at signal-to-masker ratio of 20 dB. The pattern of results for masked 85-dB SPL tones resembled that for unmasked 20-dB SL tones. The role of auditory-filter bandwidths and the existence of a compensating mechanism in the perception of across-channel synchrony and asynchrony will be discussed. [Work supported by NIH Grants R01DC03909 and R01DC010374].

4pPP11. Limiting factors in auditory discrimination of frequency ratios. Christophe N. J. Stoelinga and Robert A. Lutfi (Dept. of Communicative Disord., Auditory Behavioral Lab., Univ of Wisconsin, Madison, Wisconsin 53706)

Frequency ratios convey meaningful information for speech and many natural sounds. However, little is known regarding our ability to discriminate frequency ratios [see Fantini and Viemeister (1987). *Auditory Processing of Complex Sounds* (Lawrence, Hillsdale, NJ)]. The present study measured the relative contribution of two factors expected to limit discrimination: decision criterion and internal noise. In an adaptive, two-interval, forced-choice procedure with frequency rove, 11 highly trained listeners discriminated a change in the frequency ratio between two equal-intensity tones. The separate influence of decision criterion and internal noise was determined from scatter plots giving listeners' trial-by-trial responses as a function of the frequencies of the two tones. The plots reveal listeners to initially base their judgments on one tone alone (almost always the higher of the two) but after training to place greater reliance on both tones. Some listeners were able to achieve the optimal decision criterion of weighting both tones equally, so that the deviation from perfect performance was due only to internal noise. With continued practice over the course of many trials (at least ten 1000-trial sessions), listeners improved their performance by either improving their decision criterion or reducing internal noise. [Research supported by NIDCD grant 5R01DC006875-05.]

4pPP12. Diassociating spectral and temporal influences in an AM-QFM (amplitude modulated and quasi-frequency modulated) discrimination task. Ewa Borucki and Berg Bruce (Dept. of Cognitve Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697-5100, eborucki@uci.edu)

This study investigated temporal and spectral influences in a task used to investigate the bandwidths of phase sensitivity. Subjects discriminated amplitude modulated (AM) tones and quasi-frequency modulated (QFM) tones in a 2IFC. An adaptive threshold procedure was used to estimate modulation depth needed to make the discrimination as a function modulation frequency for a 2000-Hz carrier. Threshold functions were often nonmonotonic, with nonmonotonicities observed at higher-modulation frequencies. This is likely due to the effects of distortion products creating salient spectral cues. When stimulus duration is decreased from 200 to 50, 20, or 10 ms, thresholds for low-modulation frequencies decreased to near-chance levels, whereas thresholds in the region of nonmonotonicities were largely unaffected. The decrease in stimulus duration appears to hinder the listener's ability to use temporal cues in order to discriminate between AM and QFM, whereas spectral information derived from distortion product cues appears to be resilient. [Work supported by NSF BCS-07464003.]

4pPP13. The influence of practice on the discrimination of spectro-temporal modulation depth. Andrew T. Sabin (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60201, a-sabin@northwestern.edu), David A. Eddins (Univ. of Rochester, Rochester, NY), and Beverly A. Wright (Northwestern Univ., Evanston, IL 60201)

The pattern of sound energy spread across frequency and time (spectro-temporal modulation) is a crucial stimulus cue for sound identification. Improvements in the sensitivity to this modulation could potentially aid performance on numerous real-world tasks, yet it is unknown how practice influences this sensitivity. To investigate this issue, normal-hearing adults ($n=8$) were trained ~1 h/day for 7 days to discriminate between noises with the same spectro-temporal modulation drifting upward in audio frequency but with different modulation depths. Performance on the trained condition and on four untrained conditions was examined both before and after the training phase. Depth-discrimination thresholds improved significantly on the trained condition and on an untrained condition with a downward-drifting modulation. Discrimination thresholds also improved in subsets of listeners on untrained isolated component modulations (spectral or temporal) of the trained spectro-temporal modulation. Finally, the ability to detect the trained spectro-temporal modulation worsened significantly in proportion to the amount of improvement on the trained discrimination task. These data suggest that training on depth discrimination may be a means of improving sensitivity to spectro-temporal modulation depth in real-world stimuli, but potentially at the cost of the ability to detect these modulations. [Work supported by NIH/NIDCD]

4pPP14. Transient sex differences during development on two auditory tasks attributable to earlier maturation in males. Julia Jones Huyck and Beverly A. Wright (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60202, juliajoneshuyck@u.northwestern.edu)

Recent evidence suggests that the development of naïve performance on some auditory perceptual tasks can continue well into adolescence. Of interest here was whether there are sex differences in the maturation rate on three such tasks, temporal-interval discrimination, tone detection in forward masking, and tone detection in backward masking. To investigate this issue, performance was compared between males and females in multiple age groups on these three tasks (n per sex in each age group=9–20). There were no sex differences for backward masking. However, males reached adult-like performance earlier than females on the other two tasks: forward masking (males at ~12 years, females at ~15 years) and temporal-interval discrimination (males at 14 years, females at >14 years, on each of three

separate conditions). Surprisingly, the male advantage occurred only during adolescence. Thus, on two of the three late-developing auditory abilities examined, males matured more quickly than females, but did not differ from females either early in development or once mature performance had been reached. These results demonstrate that there can be transient sex differences during development owing to differences in maturational rate. [Work supported by NIH/NIDCD.]

4pPP15. Spectral weight analysis of commodulation masking release. Hisaaki Tabuchi and Bruce G. Berg (Dept. of Cognit. Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza. Irvine, CA 92617-5100)

Spectral weight estimates are obtained in a commodulation masking release (CMR) experiment in which masking and flanking bands each consist of five equal-intensity tones spanning 20 Hz. The masking and flanking bands are centered at 1000 and 900 Hz, respectively, and the signal is an increment in the intensity of the 1000 Hz component. The phases for the five components of the masking band are randomly sampled on each trial and are identically assigned to the components of the flanking band, yielding commodulated bands. A small, random intensity perturbation is added to each component of the stimulus, and an analysis of trial-by-trial responses provides spectral weight estimates. Spectral weights are also obtained from simulations of two multiple channel models [Hall *et al.*, J. Acoust. Soc. Am. **76**, 50–56 (1984); S. Buus, J. Acoust. Soc. Am. **78**, 1958–1965 (1985)] and one single-channel model [B. G. Berg, J. Acoust. Soc. Am. **100**, 1013–1023 (1996)]. Data for most subjects show strong support for Hall's multiple-channel model. [Work supported by NSF BCS-07464003.]

4pPP16. Relationship between physiological and psychoacoustical sensitivity to amplitude and frequency modulation. Michelle Hsieh, Craig Champlin, and Su-Hyun Jin (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, 1 University Station A1100, Austin, TX 78712, michellehsieh@mail.utexas.edu)

Various researchers have found a correlation between the auditory steady state response (ASSR) and corresponding behavioral measurements such as speech recognition scores [e.g., Dimitrijevic *et al.* (2001)]. However, relatively few studies have examined the sensitivity to small changes in amplitude and frequency modulation depths. In the present study, it was hypothesized that physiological individual differences in sensitivity to amplitude and frequency modulation depths would be reflected in corresponding psychophysical measures. Auditory steady-state responses were collected in response to amplitude- and frequency-modulated pure tone carriers (500 and 3000 Hz) in normally hearing listeners over a range of modulation depths at several different modulation rates. Participants also completed a psychophysical task in which they were asked to detect the modulated tone in a two-interval forced choice testing paradigm. The relationships between the ASSR and the psychophysical performance will be discussed.

4pPP17. A hybrid procedure for psychometric function estimation. Harisadhan Patra, Daniel L. Valente, and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131)

Threshold, defined as the stimulus level required for a predefined percent correct response, is often used to measure a listener's performance. Psychometric functions (PFs) provide better insight to the underlying decision process. Two parameters, threshold (α) and slope (β), are sufficient to define a PF. Estimation of α and β from individual trials of adaptive procedure tracks is time-consuming, costly, and dependent upon the step-size choice. A procedure has been developed and written in MATLAB, which provides stable PF parameter estimates of α and β and also their confidence intervals. The procedure consists of three stages. First, signals are presented adaptively to estimate levels corresponding to 71% and 87% correct. Second, the signal is pseudo-randomly presented at one of five fixed levels equally spaced over that range. Finally, the program implements a maximum-likelihood procedure updated after every trial to estimate signal levels corresponding to 63%, 71%, 76%, 79%, and 87% correct. After 140 trials, the program estimates α and β from the entire run and generates a PF curve that typically accounts for at least 90% of the total variance. The reliability and validity of the procedure are analyzed using simulations and observed data. [Work supported by NIH.]

4pPP18. Fundamental frequency and pitch shift discrimination. William A. Yost, Christopher A. Brown, and Farris Walling (Speech and Hearing Sci., Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287, william.yost@asu.edu)

The pitch-shift of the residue has been used as an argument for the importance of temporal fine structure in pitch perception. Discrimination of a change in fundamental frequency (F0) for harmonic complexes was compared to discriminating a change in the shifted frequency (f) in pitch-shift of the residue stimuli. Patterson and Wightman [J. Acoust. Soc. Am. **59**, (1976)] showed that a linear relationship exists between matched pitch and the frequency of the lowest spectral component for pitch shift of the residue stimuli (stimuli for which the spacing between adjacent spectral components is constant, but all components in the complex are shifted up or down in frequency). Obtaining pitch matches for pitch shift of the residue stimuli is a difficult task because of the ambiguous nature of the pitch shift of the residue in several conditions. Using the two discrimination experiments allows for an estimate of the slopes relating matched pitch to the frequency of the lowest component. Slopes estimated from the discrimination experiments will be compared to those obtained by the previous authors to determine if the discrimination experiments provide a valid way to predict the matched pitch of pitch-shift of the residue stimuli. [Research supported by NIDCD.]

4pPP19. Extraction of fundamental pitch using Euclid's algorithm implemented through iterative demodulation. Ramdas Kumaresan, Jiun-yee Li (Dept. of Elec. and Comput. Eng., Univ. of Rhode Island, Kelley Hall, Kingston, RI 02881), and Peter Cariani (Harvard Med. School, Newton, MA 02460)

Strong periodicity or "virtual" pitches are heard at the fundamentals (F0s) of harmonic complexes. As an alternative to both spectral pattern matching and autocorrelation-like analysis, we propose a novel mechanism that extracts the pitch directly from the signal by iterative demodulation and low-pass filtering. In effect the mechanism computes the greatest common divisor (gcd) of component signal harmonic numbers, using a Euclid-like iterative subtraction algorithm. For example, F0 of 560- and 320-Hz tones is 80 Hz, obtained by iteratively subtracting smaller frequencies from larger ones until they converge, e.g., $560-320=240$ Hz, $320-240=80$, $240-80=160$, and $160-80=80$. This algorithm can be adapted to compute pitches of harmonic complexes. Consider two tones, 560 and 320 Hz. As in Euclid's algorithm, the tones are iteratively multiplied and low-pass filtered to obtain the first (240-Hz), second (160-Hz), and third (80-Hz) difference tones. At the end only F0 remains. For inharmonic tones, incomplete demodulation results if there is a lower-frequency limit to the process. The mechanism converges on one pitch rather than multiple subharmonics. Such repeated multiplication and low-pass operations could conceivably be implemented in the ascending auditory pathway via phase-locked spike trains with synchrony capture, successive cross-CF convergences, low-pass modulation tunings, and cumulative decline of phase-locking at successive stations. [Work supported by AFOSR FA9550-09-1-0119.]

4pPP20. Musical context affects detection of pitch differences in tones with different spectra. Elizabeth M. O. Borchert and Andrew J. Oxenham (Dept. Psych., Univ. of Minnesota-Twin Cities, 75 East River Rd., Minneapolis, MN 55455)

Pitch plays an important role in complex auditory tasks such as listening to music and understanding speech, yet pitch comparisons are more difficult when tones have different spectra. Since tones are rarely heard in isolation, the surrounding context may help or hinder pitch comparisons. The first study presented listeners with tone pairs in isolation or immediately following a tonal context, which consisted of a portion of a descending major scale with the target as the tonic. Listeners' performance improved in the presence of the tonal context, but only when the tones within a pair differed in spectral content. In the second experiment, a variety of contexts were used, with the goal of discriminating between effects of tonal hierarchy and effects of predictability. Even with maximal predictability, presenting tone pairs after an "atonal" context consisting of notes from an octatonic scale yielded poorer performance than tone pairs in isolation. The results suggest that any advantage listeners derived from the tonal context in our first experiment

was related more to placing the tone into an over-learned tonal hierarchy than to the likelihood of the target pitch within a predictable context. [Work supported by NIH Grant R01 DC 05216.]

4pPP21. Neural timing nets for fundamental frequency (F0)-based auditory scene analysis. Peter Cariani (Dept. of Otology & Laryngology, Harvard Med. School, 629 Watertown St., Newton, MA 02460) and Ramdas Kumaresan (Univ. of Rhode Island, Kingston, RI 02881)

Neural timing nets are idealized networks of delay lines, coincidence detectors, and adaptive processing elements that operate in the time domain on temporally coded signals to compare, extract, and separate auditory objects [Cariani, Neural Networks **14**, 737–753 (2001); J. New Music Res. **30**, 107–135; IEEE Trans. Neural Netw. **15**(5) (2004)]. Timing nets constitute an alternative, potential mode of neural signal processing in which information resides in neural signals rather than in patterns of activated elements. Recurrent timing nets with delay loops act as dense arrays of recursive, comb-like filters to effect a period-by-period analysis that builds up and separates component invariant time patterns with different fundamentals (F0s). Using both linear and nonlinear processing rules, the latter were used to process and separate synthetic double vowels, running speech, and polyphonic musical excerpts, with varying results. Relations to, and combinations with, processing strategies based on autocorrelation and all-order interspike intervals, adaptive comb filtering, correlogram duplex analysis, cancellation, Fourier zero-crossings, and demodulation, with and without prior bandpass filtering, are discussed. Bottom-up/top-down mechanisms for dynamic facilitation of lower level temporal processing loops are also considered. [Work supported by AFOSR FA9550-09-1-0119.]

4pPP22. The effect of stimulus context on cortical measures of pitch processing using functional magnetic resonance imaging. Daphne García (MRC Inst. of Hearing Res., Univ. Park, Nottingham, United Kingdom, daphne@ihr.mrc.ac.uk), Christopher Plack (Univ. of Manchester, Manchester, United Kingdom), and Deborah Hall (Nottingham Trent Univ., Nottingham, United Kingdom)

Many different paradigms and pitch-evoking stimuli have been used to study pitch. A growing body of neurophysiological evidence shows that cortical responses are sensitive to the context from which stimuli are presented. In this human fMRI study, we tested the hypothesis that the stimulus context influences the pattern of pitch-related auditory cortical activation. Fifteen listeners participated in a blocked design experiment. A diotic harmonic complex tone and a dichotic Huggins pitch stimulus were presented within either a noise or silent context and activation was contrasted with matched noise control conditions. Results revealed significant main effects of both context (noise/silence) and pitch stimulus (diotic/dichotic) ($p < 0.05$, corrected). While the response to context was primarily localized in primary auditory cortex, the response to pitch was more posterior. The context significantly modulated the pitch response, especially in subregions of planum temporal; non-primary auditory cortex. While the response to the diotic pitch was greater than the dichotic pitch, the overall results were broadly comparable. We therefore conclude that if the feature specificity of the pitch-related response is to be inferred from fMRI data, future studies should include careful controls for stimulus context.

4pPP23. Vocal pitch regulation depends on the baseline voice F0 feedback: An ERP study for investigating the role of auditory feedback for voice pitch error correction. Roozbeh Behroozmand and Charles Larson (Dept. of Commun. Sci. and Disord., Speech Physio. Lab., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, r-behroozmand @northwestern.edu)

Voice production and control require neural interactions between the vocal motor and auditory mechanisms. The comparison between the incoming auditory feedback and the predicted sensory input (efference copies) from a self-produced vocalization allows the detection of feedback mismatches for voice error detection and correction. However, the sensitivity of the audio-vocal system for feedback mismatch detection seems to depend on the extent of feedback deviation from the predicted vocal output. The present study investigated this effect by examining event-related potentials (ERPs) in response to a +100-cent voice feedback pitch perturbation stimulus while the extent of pre-stimulus (baseline) feedback pitch deviation was randomly manipulated at 0, 50, 100, 200, and 400 cents. Results showed that the neural responses to +100-cent pitch-shift stimuli grew systematically larger as the extent of pre-stimulus baseline pitch deviation became smaller. This finding suggests that the extent of disparity between the predicted and incoming sensory feedback of self-produced voice can affect the neural tuning processes that adjust the sensitivity of the audio-vocal mechanisms for voice pitch error detection and correction. Lower sensitivity to larger feedback mismatches may imply robustness against the disruptive effect of highly deviant or externally generated sounds during vocal production and control.

4pPP24. Molecular analysis of the effect of onset asynchrony in the identification of a rudimentary sound source in noise. Robert A. Lutfi, Ching-Ju Liu, and Christophe N. J. Stoelinga (Dept. of Communicative Disord., Auditory Behavioral Res. Lab., Univ. of Wisconsin, 1975 Willow Dr., Madison, WI 53706, ralutfi@wisc.edu)

The threshold for detecting a target in noise is often greater when the target is gated on simultaneously with the noise than when it is gated on after some delay [Zwicker, E. (**1965**). J. Acoust. Soc. Am. **37**, 653–663]. One explanation is that the perceptual principle of grouping causes the target and noise with simultaneous onsets to be perceived as a single sound source. This idea was tested for a rudimentary sound source using perturbation analysis. In a two-interval, forced-choice procedure listeners identified as target the impact sound produced by the larger of two stretched membranes. The noise on each presentation was the impact sound of a variable-sized plate. Grouping predicts that the decision weights on the noise should be positive when target and noise have simultaneous onsets, but that they should approach zero when target and noise are gated on asynchronously. This prediction was confirmed when the noise preceded the target by a fixed interval (100 ms), but not when it followed the target by the same interval or when either interval was selected at random on each presentation (100 or −100). [Research supported by NIDCD Grant 5R01DC006875-05.]

4pPP25. Target enhancement and noise cancellation in the identification of rudimentary sound sources in noise. Robert A. Lutfi, Ching-Ju Liu, and Christophe N.J. Stoelinga (Dept. of Communicative Disord., Auditory Behavioral Res. Lab., Univ. of Wisconsin, 1975 Willow Dr., Madison, WI 53705, ralutfi@wisc.edu)

The identification of targets in noise is believed to entail two processes, one that acts to enhance the target and the other to cancel the noise [Durlach *et al.* (2003). J. Acoust. Soc. Am. 2984–2987]. The relative contribution of these processes in the identification of rudimentary sound sources is unknown. In the present study, perturbation analysis was used to determine the relative contribution in terms of the sign and magnitude of listener decision weights on the noise. In a two-interval, forced-choice procedure, listeners identified as target the impact sound produced by the larger of two stretched membranes. The noise on each presentation was the impact sound of a variable-sized plate. For four of five listeners showing significant interference, the weights were positive indicating enhancement; for the remaining listeners, they were negative indicating cancellation. In a second condition, the target was the membrane hit with harder force, and the noise was a plate hit with variable force. The noise weights for all listeners in this condition indicated cancellation. The results are consistent with an interpretation in which noise cancellation is the predominant strategy for identifying targets unless it is precluded by uncertainty regarding the spectral properties of the noise. [Research supported by NIDCD Grant No. 5R01DC006875-05].

4pPP26. The effect of binaural coherence on envelope interaural time difference thresholds. Jessica J. M. Monaghan, Katrin Krumbholz, and Bernhard U. Seeber (MRC Inst. of Hearing Res., Nottingham NG7 2RD, United Kingdom, jessica@ihr.mrc.ac.uk)

Room reflections alter the envelope of sounds differently at both ears, reducing binaural coherence. Experiment 1 measured interaural time difference (ITD) discrimination thresholds for broadband and transposed speech. A sentence was convolved with binaural room impulse responses for different source-receiver distances, and the envelopes extracted and multiplied with 4-kHz tones. Preliminary results indicate that envelope ITD thresholds increased from <100 μs without reverberation to >700 μs at four times the reverberation radius. ITD thresholds for unprocessed speech were more robust to the addition of reverberation, only rising above 350 μs at the same distance. To investigate whether reverberation is detrimental owing to re-

duction in binaural coherence, stimulus envelopes were created by temporally jittering raised cosine pulses around a 10-ms separation. Bilaterally independent jittering allowed variation in coherence while minimizing change in other envelope parameters. Preliminary results show ITD thresholds <100 μs for coherent envelopes, increasing to >700 μs for coherences of 0.6, a value consistent with the room simulations above. Envelope coherence strongly affects ITD discrimination, suggesting that ITDs extracted from high-frequency channels may not provide useful information in many realistic situations. This has implications for bilateral cochlear implant users, as current devices provide ITDs only in envelopes.

4pPP27. Tonotopic gradients in neural interaural time difference processing: A modeling study of the medial superior olive. Yoojin Chung (Eaton-Peabody Lab., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114, yoojin_chung@meei.harvard.edu) and H. Steven Colburn (Boston Univ., Boston, MA 02215)

The interaural time difference (ITD) is the primary cue for sound localization; yet many basic questions about the biophysical mechanisms remain. In the mammalian auditory system, neurons in the medial superior olive (MSO) are tuned to ITD as well as frequency. In this study, the effects of key parameters in MSO models, such as the membrane time constants and the number and strength of synaptic inputs on ITD sensitivity to pure tones, were explored with particular attention to the best-frequency dependence of these parameters. Results show that ITD tuning is dependent on neuron membrane characteristics (i.e., the shape and time constant of the membrane response), on the strength of individual synaptic inputs, and on stimulus properties such as frequency and intensity. Models with slow response times and weak synaptic inputs show good ITD sensitivity for low-frequency tones, whereas models with fast response times and strong synaptic inputs exhibit good ITD tuning for high-frequency tones. This dependence of ITD sensitivity on membrane properties in the model suggests that parameters important for ITD tuning depend on the best frequency of the neurons, in contrast to the view that the binaural mechanism is homogeneous along the tonotopic axis.

4pPP28. Level-dependent changes in perception of temporal envelope cues. Xin Wang, Jayne B. Ahlstrom, Amy R. Horwitz, and Judy R. Dubno (Dept. of Otolaryngol.-Head and Neck Surgery, Medical Univ. of South Carolina, 135 Rutledge Ave., MSC 550, Charleston, SC 29425-5500, wanxi@musc.edu)

Level-dependent changes in speech recognition may reveal effects of basilar-membrane nonlinearities on temporal envelope fluctuations. It is hypothesized that, as a result of the compressive effects of the active cochlear mechanism, the "effective" magnitude of speech envelope fluctuations will be reduced as speech level increases from lower (more linear) to conversational (more compressive) regions. With further increases from conversational levels (to a more linear region), temporal envelope fluctuations will become more pronounced. Accordingly, speech recognition will be maximized at conversational levels due to the optimal "flattened" envelope and then decline at lower or higher levels. To test these assumptions, speech recognition scores were measured as a function of level for adults with normal hearing. Speech stimuli were spectrally degraded using "noise vocoder" processing so that perceptual effects of modifications to the speech temporal envelope can be revealed. As vocoded speech level increased, background noise level also increased, maintaining a fixed signal-to-noise ratio to minimize sensation-level effects on speech recognition scores. Discussion will focus on level-dependent effects for different speech stimuli and the role of nonlinearities on perception of temporal envelope cues. [Work supported by NIH/NIDCD]

4pPP29. The use of confusion patterns to evaluate the neural basis for concurrent vowel identification. Ananthakrishna Chintanpalli (Weldon School of Biomedical Eng., Purdue Univ., 206 S. Martin Jischke Dr., West Lafayette, IN 47907) and Michael G. Heinz (Purdue Univ., West Lafayette, IN 47907)

Perceptual studies of concurrent vowel identification suggest that listeners with normal hearing (NH) are better able to use differences in fundamental frequency (F0) than listeners with sensorineural hearing loss (SNHL). However, the neural basis for this difference remains unknown. The present study sought to validate a neural model of concurrent vowel identification based on specific confusion patterns made by NH listeners. A standard set of five vowels was used. In each concurrent vowel pair, vowel 1 had F0 = 100 Hz, while the F0 of vowel 2 was varied between 100–126 Hz (4 semitones). NH listeners made similar confusions across all F0 differences, with a reduction in occurrence as F0 difference increased. F0 benefit varied significantly across concurrent vowel pairs. A phenomenological auditory-nerve model was cascaded with F0 segregation algorithms to predict the perceptual observations. Neural predictions showed similar confusion patterns to the perceptual data for many (but not all) concurrent vowel pairs. Validating a NH physiological model with specific confusion patterns will allow the model to be used to predict effects of specific physiological changes related to SNHL, which may be useful for improving cochlear-implant and hearing-aid signal processing strategies. [Work supported by Purdue University.]

4pPP30. Auditory channel weights for consonant recognition in normal-hearing listeners. Frederic Apoux and Eric W. Healy (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210, fred.apoux@gmail.com)

The present study evaluated the relative contribution of various regions of the frequency spectrum to consonant recognition in normal-hearing listeners. The method used in this study was specifically designed to provide an estimate of the importance of each band (i) consistent with the frequency resolution of the auditory system and (ii) irrespective of the location of information elsewhere in the spectrum. Speech stimuli were divided into 30 adjacent bands with each band corresponding approximately to the bandwidth of an auditory filter. Listeners were presented with a subset of bands to avoid ceiling effects. The importance of each band was derived from the drop in performance observed when that particular band was omitted. The spectral location of the bands was always chosen randomly except for the one band whose importance was being assessed. The results indicated a fairly homogeneous contribution of all 30 bands to consonant recognition (i.e., a flat auditory channel importance function) with only the five lowest bands (below 300 Hz) having lesser weight. In contrast, additional analyses revealed a non-uniform contribution of the bands to the transmission of voicing, manner and place of articulation. [Work supported by NIDCD.]

4pPP31. Perception of temporally interrupted speech: Effects of two concurrent gating rates on intelligibility. Valeriy Shafiro, Stanley Sheft, and Robert Risley (Comm. Dis. Sci., Rush Univ. Med. Cntr., 1015 AAC, 600 S. Paulina St., Chicago, IL 60612, valeriy_shafiro@rush.edu)

Perception of temporally interrupted speech was investigated with either one or two concurrent square-wave gating functions. Sentences were interrupted at a single rate of 1, 2, 4, 8, 16, 24, or 32 Hz with a 50% duty cycle. For each rate between 1 to 8 Hz, stimuli were additionally gated at a secondary rate ranging from twice the primary rate up to 32 Hz. The secondary gating rate thus interrupted only the segments of speech which were left intact after application of the primary gating function to reduce by half the remaining speech content independent of secondary rate. With a single gating function, intelligibility scores increased with rate, reaching perfect accuracy at 8 Hz. Application of the second gating function led to a decrease in intelligibility which was greatest when the secondary rate was twice the primary rate, a condition equivalent to a 25% duty cycle of the primary rate. Further increase in the secondary rate generally led to an improvement in intelligibility, although never to the level achieved with the primary rate alone. The incomplete recovery of intelligibility scores at fast secondary gating rates suggest disruption of separate perceptual processes by each of the two concurrent interruption rates.

4pPP32. Noise robust representation of speech in the primary auditory cortex. Nima Mesgarani (Johns Hopkins Univ., Baltimore, MD, mnima@umd.edu), Stephen David, Jonathan Fritz, and Shihab Shamma (Univ. of Maryland, College Park, MD)

It is well known that humans can robustly perceive phonemes despite substantial variability across speakers, context and natural distortions. This study examines the responses of neurons in primary auditory cortex (A1) to phonetically labeled speech stimuli in clean, additive noise and reverberant conditions. Using a linear decoder [Bialek (1991)] to reconstruct the input stimulus spectrogram from the population response, we observed that

spectrograms reconstructed from the neural responses to noisy speech were closer to the original clean spectrograms than to the noisy ones. This indicates that sound representations in A1 serve to enhance information about natural speech signals relative to noise, thus extracting signal from noise. Examining the average reconstructed phoneme spectrograms in clean and noisy speech revealed a remarkable robustness in the encoding of features important for discrimination of different phonemes. In addition, it was found that the strict linear spectro-temporal receptive field (STRF) model of A1 neurons is insufficient to explain the noise robustness observed in the neural data. However, when a non-linear synaptic depression is integrated into the inputs for the STRF model, the noise was reduced in the reconstructed spectrograms similar to what observed with the actual neural data.

THURSDAY AFTERNOON, 22 APRIL 2010 DOVER A/B, 2:00 TO 4:45 P.M.

Session 4pSA

Structural Acoustics and Vibration: Applications of Structural Acoustics and Vibrations II

Joel M. Garrelick, Chair
Applied Physical Sciences, 49 Waltham St., Ste. 4, Lexington, MA 02141

Contributed Papers

2:00

4pSA1. Wave propagation model for acoustic evaluation of polymeric thin films. Hyeong Sick Ju (Graduate Program in Acoust., The Pennsylvania State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, huj110@psu.edu) and Bernhard R. Tittmann (The Penn State Univ., University Park, PA 16802)

Polymeric thin films are fabricated on crystalline wafers to serve as a photoresist in lithography processes of semiconductors or MEMS. Essentially these film structures are viscoelastic layers on anisotropic substrates. Acoustic evaluation for these material structures requires an appropriate wave propagation model. This paper presents a leaky surface acoustic wave (LSAW) model for the evaluation of polymeric film integrity. Scanning acoustic microscopy (SAM) operating at relative high-frequency above 100 MHz is utilized to measure the LSAW velocity. For the use of SAM, the model modifies the pre-existing anisotropic layered model by employing hysteretic absorption in viscoelastic polymer and water-loading. The model produces the mean reflectance function and mean dispersion curve, which are specialized for spherical acoustic lenses. The predicted mean dispersion shows good agreement with LSAW velocities measured at several frequencies with spherical lenses.

2:15

4pSA2. Active vibration control modules for damping, compensation, measurement and dynamic testing. Vyacheslav M. Ryaboy (Newport Corp., 1791 Deere Ave., Irvine, CA 92606, vyacheslav.ryaboy@newport.com)

The active vibration control module was introduced previously as part of active vibration damping system for optical tables and other precision vibration isolated platforms. This paper describes steps to expand the application of those modules to other tasks, namely, compensation of forced vibration in local areas and dynamic testing of structures. Vibration damping of most significant structural modes had been achieved using a small number of properly placed active dampers. This did not affect forced tonal vibration. Current state of the art does not offer a practically feasible way to suppress all vibration, forced and normal, over the total table surface. However, by placing vibration control modules around a local area of the table supporting a vibration sensitive device, it is possible to abate forced tonal vibration in this area. Feedback control for vibration compensation will be discussed along with experiments demonstrating stable concerted work of several vibration control modules. In application to dynamic testing, the actuator is excited by white noise, and the sensor signal is processed to calculate the dynamic compliance. The test data show that the vibration control modules can be used to measure dynamic compliance with precision comparable to that of dedicated vibration measurement systems.

2:30

4pSA3. When is a system random? Richard H. Lyon (60 Prentiss Ln., Belmont, MA 02478)

Statistical energy analysis (SEA) is a method for estimating structural acoustic transfer functions based on a statistical model of the system. The SEA model assumes random noise excitation in frequency bands and an underlying ensemble of "similar" systems. However, since the results of a SEA calculation are nearly always applied to a single system, there is a natural question of in what sense a single system can be random. A particular structural acoustics system is considered that displays both deterministic and random behavior with a distinct transition between the two. System parameters (damping, modal density, and coupling factors) are examined to see how they might be controlling the transition from deterministic to random.

2:45

4pSA4. Novel lightweight vibration absorbers for marine structures. Ryan L. Harne and Chris R. Fuller (Virginia Tech Vib. and Acoust. Labs., 131 Durham 0238, Blacksburg, VA, 24061, rharne@vt.edu)

When combined with attached motors and rotating machinery, the lightly damped, thick plating required in maritime applications becomes a broadband noise and vibration control problem. A typical solution is to adhere heavy and dense damping materials for dissipation of the plate vibrational energy. In order to attenuate low frequencies, significant mass must be added to the structure. This paper will review the development of two, new passive treatments intended to resolve this issue. HG blankets are constructed using small masses embedded into poroelastic material. Together with the inherent stiffness of the poroelastic material, the masses become embedded mass-spring dampers and their presence is found to notably increase the low-frequency transmission loss of the host material. DVAs are compact vibration absorbers that distribute continuous mass and spring elements over the surface while generating ample reactive damping at low frequencies. This paper will overview the concepts and development of adapting DVAs and HG blankets for use on heavy plate structures, their testing for broadband control performance, as well as their versatility for thinner panels. A comparison with a conventional, marine noise control treatment will be considered. [This work was supported by Northrop Grumman Shipbuilding-Newport News.]

3:00

4pSA5. Bloch response of framed wings with upper and lower skins. R. Martinez and M. Eash (Alion Corp., 84 Sherman St., Cambridge, MA 02143)

Aircraft wings are typically composed of spanwise-periodic framing elements running from leading to trailing edge. Each frame is connected to the wing's upper and lower skins through continuity of motions among frames

and skins. This paper presents calculations of a generic wing's low-wavenumber Bloch response in global flexure per the following physical content: (1) the frames are mass controlled both in vertical translation and rotation; (2) the skins are plate-like, and respond dynamically to normal and compressional shears against each of the frames; (3) the two plates bend symmetrically with respect to the normal component of the virtual foces driving them, and antisymmetrically in compression with respect to their tangent component. The need for compressional forces and a corresponding in-plane response for each of the wing's two skins stems from the theory's lack of restrictions on frame height, which for our thick wing implies a neutral axis for global flexure that is far from the system's uppermost and lowermost material points. Our calculations showcase the effects of the skins' compression on the system's low-wavenumber response, as well as how compression defines the structure's low-frequency limit for a stiffness-controlled effective medium rendered anisotropic by its chordwise-running frames.

3:15—3:30 Break

3:30

4pSA6. Noise and vibration transmission reduction using multi-element multi-path structures. Cassidy Palas and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Hudson Hall, Durham, NC 27708, sid.palas@duke.edu)

Principles of structural acoustics are utilized in novel ways to cancel the transmission of sound and vibration through multi-element flexible barriers. Configurations analyzed include two and three layered plates with elastic interconnections. The substructures have different wave propagation properties and boundary conditions. Both continuous and periodic discrete elastic coupling methods are examined. This research demonstrates that flexibility and controlled resonant behavior can be used to block vibration and sound transmission, even with low structural damping. The main strategies utilized are structural wave cutoff with multi-element multi-path (MEMP) structures and relative phase changes due to boundary reflections. Examples of acoustic transmission loss through panel barriers are presented, and the potential advantages and possible shortcomings of the approach are evaluated. Practical configurations for layered sound reduction materials include designs allowing multiple substructural plates to produce radiation on a given surface, leading to net cancellations of transmitted sound in certain frequency ranges. Experimental results show vibration transmission reduction for several configurations. The work has particular application to the reduction in vehicle interior noise and addresses the need for good acoustic performance of lighter weight flexible structures.

3:45

4pSA7. An improved formulation for predicting low-frequency noise transmitted through double-pane windows. Dayi Ou, Cheuk Ming Mak (Dept. of Bldg. Services Eng., The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong, 08900950r@polyu.edu.hk), and Kai Ming Li (Purdue Univ., West Lafayette, IN 47097-2031)

With superior sound insulation properties over a single-panel configuration, a double-panel structure with the presence of a cavity has found a wide range of applications for sound insulation. A classical method for combining a finite element method (FEM) with a boundary element method (BEM) is used to examine the transmission of low-frequency noise through double-pane windows in the present study. The technique of component mode synthesis is applied to adjust the stiffness matrix in the FEM formulation in order to examine the effects of elastic boundary conditions on the sound transmission through these structures. However, the Green function for predicting sound propagation in a rectangular long enclosure is used in favor of the free-field Green function for predicting the pressure inside the cavity of the window panes. The predicted pressure in the cavity is then coupled with the FEM formulation for the window panes and the BEM formulations for the sound fields at the outer surfaces of window panes. A parametric study is conducted systematically to allow a detailed examination for the characteristics of sound insulation of a double-pane window at different frequency bands especially for the low-frequency components.

4:00

4pSA8. Acoustic signatures of partial electric discharges in different thicknesses of Kapton. Daniel P. Hanley (US Naval Acad., P.O. 12571, Annapolis, MD 21412, m102592@usna.edu) and Edward J. Tucholski (US Naval Acad., Annapolis, MD 21401)

Thin polymer films in the presence of high electric fields undergo partial discharge and have characteristic acoustic emissions. It is hoped that studying these acoustic signals can aid in anticipating failure of the films, thereby providing a tripwire to reduce the electric voltage in high-energy capacitor applications before failure actually occurs and the capacitor is permanently damaged. This study compared the acoustic emission from a variety of thicknesses of Kapton film between 7 and 55 μm. A laser Doppler vibrometer with a frequency response from 0–22 kHz was used to study surface vibrations of a gold coated polymer sample as voltage was raised at a controlled rate of 500 V/s from 0 V to material failure. Emissions from partial discharges prior to failure are studied. The results of these tests demonstrate the relationship between characteristic frequencies for the acoustic emission and the polymer thickness.

4:15

4pSA9. Seismic surface wave method for near surface soil exploration. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of MS, Univ. MS 38677, zhiqulu@olemiss.edu)

There are many applications that require the information of near surface soil properties. The related areas include agricultural land management, levee/dam evaluation, landmines/UXO/tunnel detection, battle field condition assessment, and site foundation characterization, to name a few. To obtain soil properties in a non-invasive manner, a multi-channel analysis of surface wave (MASW) method based on laser-Doppler vibrometer was developed recently and reported in 157th ASA Meeting, Portland, OR. This talk will present the latest development. In particular, two methods of determination of the dispersive curve, i.e., the phase slope method and the two dimensional FFT transformation method, will be discussed. The emphasis will be given to demonstrate the capabilities of the two methods in identifying the fundamental mode and high modes of Rayleigh waves. An inversion algorithm using the measured dispersive curve can back calculate the soil profile, i.e., the shear wave velocity as a function of depth. Several case studies of the MASW method will be addressed.

4:30

4pSA10. Fault classification for rotating machines using neural network. Hyungseob Han and Uipil Chong (Dept. of Comput. Eng. and Information Technol., Univ. of Ulsan, Ulsan, Korea, overhs@naver.com)

When the rotating machines in the plants malfunction during operating, they can cause huge economic losses and many casualties. For these reasons, fault detection and diagnosis of rotating machines have become very important issues. This paper proposes a system to detect and diagnose abnormal states for induction motors. Through an effective combination of wavelet transform and neural network for measured vibration signals from the motors, successful fault detection and diagnosis can be achieved. This system is divided into two parts: analysis and classification. In the analysis part, vibration signals are divided into eight subbands by using wavelet transform. The most significant features of the signals are shown in the lowest-frequency band. For an efficient classification in the neural network, input samples chosen in that band were minimized. In the classification part, through one of the representative techniques, multi-layer perceptron, all kinds of vibration signals are trained and tested. From the experimental results, the proposed system perfectly classified input signals into each fault case. Furthermore, since it does not need reference data for classification, it can perform very quickly and be implemented to a real-time system.

Session 4pSCa

Speech Communication: Speech for Tracking Human Health State, Performance, and Emotional State I

Suzanne E. Boyce, Chair
Univ. of Cincinnati, Dept. of Communication Disorders, Cincinnati, OH 45267-0379

Chair's Introduction—1:00

Invited Papers

1:05

4pSCa1. Speech analysis in accident investigation. Malcolm Brenner (Natl. Transportation Safety Board, 490 L'Enfant Plaza SW, Washington, DC 20594, brennem@ntsb.gov)

The NTSB investigates major transportation accidents in the United States to make recommendations to prevent their recurrence. The Safety Board also examines new technologies that might assist investigations. In conducting its work, the NTSB has found speech analysis a useful new technology for providing secondary evidence on operator state for issues such as psychological stress, alcohol impairment, physical straining, and hypoxia. This talk provides examples from two investigations: the grounding/oil spill of the Exxon Valdez tanker and the crash of a Boeing 737 airliner at Pittsburgh.

1:25

4pSCa2. Common voice measures as indicators of fatigue. Cynthia M. LaJambe (201 Transportation Res. Bldg., Penn State Univ., University Park, PA 16802, cml149@psu.edu), Frederick M. Brown (Penn State Univ., University Park, PA, 16802), Rebecca M. Reichardt (Towson Univ., Towson, MD, 21252), Malcolm Brenner (Natl. Transportation Safety Board, Washington, DC, 20594), and Robert A. Prosek (Penn State Univ., University Park, PA, 16802)

Unobtrusive, economical, and readily accessible fatigue-monitoring technologies are needed especially in transportation, military operations, and security industries. Voice analysis is compatible with operational settings, given its minimal interference with hands-on work duties. Controlled laboratory studies are underway to establish the sensitivity of this fatigue-monitoring method. A recent study evaluated sleep-deprivation consequences on basic voice attributes using multiple speech tasks. Twenty-six native English-speaking 18-26 year-old subjects were screened for physical and psychological problems. Several sleep/wake cycles were monitored with actigraphy prior to laboratory participation. Vocal measures were compared between 13 speakers sleep deprived for 36 hours and 13 non-sleep-deprived controls. In the laboratory, speech was recorded during baseline sessions and on the following day. Group differences varied by speech task and vocal measure, with more sleep-deprivation sensitivity found, for example, in speaking rate as compared to fundamental frequency. Fatigue-related changes in vocal measures were associated with decrements in psychomotor reaction times and cognitive performance. Results are compared with previous studies relating psycho-physiological states to basic voice measures. Design considerations are discussed.

1:45

4pSCa3. Speaker assessment: The impact of environment on speech systems and individuals. John H. L. Hansen (Dept. of Elec. Eng., CRSS: Ctr. Robust Speech Systems, Univ. of Texas at Dallas, Richardson, TX 75083, john.hansen@utdallas.edu)

Assessing speaker variability is critical in developing a scientific understanding or model of human speech production. The environmental context plays a significant role in how this variability plays out. In this study, recent findings are presented on the variability of speech production due to environmental factors that influence man-machine interaction as well as human-to-human interaction. Speech produced under task stress, emotional stress, and background noise (resulting in Lombard effect) all cause speech production changes. This impacts both speech processing algorithms intended for speech recognition/technology and human-to-human interaction. Specifically, two speech production domains are briefly considered including Part-1: speech production under varying types/levels of background noise and how this produces flavors of Lombard effect and impacts speaker recognition systems, and Part-2: assessing the stress/emotional state of parents/care-givers in quantifying the language learning exposure of children (ages 10–36 months). In Part-1: the UT-Scope corpus is employed with speech from 30 subjects (19M,11F) for analysis of duration and spectral tilt as well as developing an automatic Lombard effect classification scheme which is incorporated into speaker recognition. Next, Part-2: considers how neutral versus stressed/emotional state of adults impacts conversational turns and adult word-count in a child language learning environment (20 child-parent interactions).

2:05

4pSCa4. Intelligibility of speech produced under sleep-deprivation conditions. Suzanne Boyce (Dept. of Comm. Sci. Disord., Univ. of Cincinnati, 3202 Eden Ave., Cincinnati, OH 45267-0379, boycese@ucmail.uc.edu), Joel MacAuslan (S.T.A.R. Corp., Bedford, MA 01730), Sandra Combs, and Alexandra Blood (Univ. of Cincinnati, Cincinnati, OH 45267-0379)

In previous work, we applied the technique of acoustic landmark detection to speech produced under rested vs sleep-deprived conditions. We found significant differences for both the number and pattern of landmarks detected. In a parallel study we found significant differences in clear vs conversational styles of speech. While it has been shown by a number of investigators that clear and conversational speech styles differ in the degree to which intelligibility to listeners is preserved in noise, it is not clear whether the articulatory changes found in sleep-deprived speech affect the ability of listeners to understand what is said, especially in noise. In this paper, we present the results of a study in which normal-hearing listeners are asked to transcribe speech presented with and without background noise. Similar levels of background noise have been shown to reduce speech intelligibility for clear vs conversational speech presented to normal listeners. The speech in this study was produced under rested and sleep-deprived conditions. Results will be compared to effects of clear vs conversational speech presented in noise and in quiet.

2:25

4pSCa5. On the acoustics of emotion in speech: Desperately seeking a standard. Bjoern Schuller (Inst. for Human-Machine Commun., Technische Universitaet Muenchen, D-80333 Muenchen, Germany, schuller@tum.de)

Researchers concerned with automatic recognition of human emotion in speech have proposed a considerable variety of segmental and supra-segmental acoustic descriptors. These range from prosodic characteristics to voice quality to acoustic correlates of articulation and represent unequal degrees of perceptual elaboration. Recently, evidence has been reported from first comparisons on multiple speech databases that spectral and cepstral characteristics have the greatest potential for the task [B. Schuller *et al.*, Linguistic Insights **97**, 285–307 (2009)]. Yet, novel acoustic correlates are constantly proposed, as the question of the optimal representation remains disputed. The task of evaluating suggested correlates is non-trivial, as no agreed "standard" set and method of assessment exists, and inter-corpus substantiation is usually lacking. Such substantiation is particularly difficult owing to the divergence of models employed for the ground-truth description of emotion. To ease this challenge, using the potency-arousal-valence space as the predominant means for mapping information stemming from diverse speech resources, including acted and spontaneous speech with variable and fixed phonetic content on well-defined binary tasks is proposed. Among the various options for automatic classification, a method combining static and dynamic features representing pitch, intensity, duration, voice quality, and cepstral attributes is recommended.

2:45

4pSCa6. The detection of stress, emotion, and deception from speech: The intersection of phonetics, policy, and politics. James Harnsberger (Dept. Comm. Disord., Univ. of Florida, 68 Dauer Hall, Gainesville, FL 32611)

While prior research on the detection of stress, emotion, and deception from speech and language has shown limited progress, this has not prevented the marketing of commercial devices that purport to detect these states to a variety of customers, such as law enforcement agencies, the military, intelligence agencies, homeland security, and insurance companies. For the major products currently on the market, all independent studies to date have failed to verify their efficacy with a wide range of speech materials collected under various experimental conditions, ranging from laboratory studies with carefully controlled speech to "mock crimes" to speech produced under realistic levels of jeopardy. This literature (including two studies by the author) will be reviewed and discussed in terms of how their experimental design and results are shaped and used in policy debates by private manufacturers, elected officials and their staffers, academic researchers, and others.

3:05—3:20 Break

3:20

4pSCa7. Automatic methods to monitor the speech of Parkinson's patients with deep brain stimulators. Craig van Horne (Caritas Neurosurgery, 736 Cambridge St., CCP 8, Brighton, MA 02135), Karen Chenausky, Joel MacAuslan (STAR Corp., Bedford, MA 01730), Carla Massari, and Marianna McCormick (Caritas-St. Elizabeth's, Brighton, MA, 02135)

Parkinson's disease (PD) is a neurodegenerative disease causing hypokinetic dysarthria, associated with "blurred" or underarticulated speech, imprecise consonants, and, sometimes, irregular syllable trains. Within the past decade, deep brain stimulation (DBS) of the subthalamic nucleus (STN) has provided substantial benefit to PD patients. DBS treatment has largely been directed toward the motoric features of PD: bradykinesia, rigidity, and tremor, but its effects on speech vary. The speech of PD patients receiving DBS treatment, with or without accompanying medical therapy, was analyzed for rate (syllables per second), regularity (relative deviation of syllable length), stop consonant spirantization (a measure of stop consonant precision), vowel ratio (length of vowel to length of syllable), and other features using automatic routines written specifically for the purpose. Patients' speech is more variable on DBS stimulation than on medication or no treatment. It is possible to find a combination of DBS settings for each patient that relieves their motor symptoms and returns their speech to normal. These findings suggest that it is possible to improve speech along with the general motor symptoms of PD. Furthermore, automatic analyses show promise as sources of feedback for neurologists to use in optimizing DBS settings for speech.

4pSCa8. Impact of cognitive load and frustration on drivers' speech. Hynek Bořil (Erik Jonsson School of Eng. and Comp. Sci., The Univ. of Texas at Dallas, 2601 N. Floyd Rd. 75080, Richardson, TX 75083-0688, hynek@utdallas.edu), Tristan Kleinschmidt (Speech and Audio Res. Lab., Queensland Univ. of Technol., GPO Box 2434, Brisbane, Queensland 4001, Australia), Pinar Boyraz, and John H. L. Hansen (The Univ. of Texas at Dallas, Richardson, TX 75083-0688)

Secondary tasks such as cell phone calls or interaction with automated speech dialog systems (SDSs) increase the driver's cognitive load as well as the probability of driving errors. This study analyzes speech production variations due to cognitive load and emotional state of drivers in real driving conditions. Speech samples were acquired from 24 female and 17 male subjects (approximately 8.5 h of data) while talking to a co-driver and communicating with two automated call centers, with emotional states (neutral, negative) and the number of necessary SDS query repetitions also labeled. A consistent shift in a number of speech production parameters (pitch, first format center frequency, spectral center of gravity, spectral energy spread, and duration of voiced segments) was observed when comparing SDS interaction against co-driver interaction; further increases were observed when considering negative emotion segments and the number of requested SDS query repetitions. A mel frequency cepstral coefficient based Gaussian mixture classifier trained on 10 male and 10 female sessions provided 91% accuracy in the open test set task of distinguishing co-driver interactions from SDS interactions, suggesting—together with the acoustic analysis—that it is possible to monitor the level of driver distraction directly from their speech.

THURSDAY AFTERNOON, 22 APRIL 2010 GRAND BALLROOM V, 4:05 TO 5:05 P.M.

Session 4pSCb

Speech Communication: Speech for Tracking Human Health State, Performance, and Emotional State II (Poster Session)

Suzanne E. Boyce, Chair
Univ. of Cincinnati, Dept. of Communication Disorders, Cincinnati, OH 45267-0379

Contributed Papers

All posters will be on display and all authors will be at their posters from 4:05 p.m. to 5:05 p.m.

4pSCb1. Using temporal cycles in spontaneous speech to quantify linguistic impairments in patients with neurodegenerative disorders. Eden Kaiser (Prog. in Linguist., Univ. of Minnesota, 214 Nolte Ctr., 315 Pillsbury Dr. SE, Minneapolis, MN 55455, kaise113@umn.edu), Serguei V. S. Pakhomov, Angela K. Birnbaum, Daniel Boley (Univ. of Minnesota, Minneapolis, MN 55455), and David S. Knopman, (Mayo Clinic, Rochester, MN 55905)

Fronto-temporal lobar degeneration (FTLD) is a form of dementia which may manifest through symptoms similar to Alzheimer's, including language-specific impairment. Linguistic manifestations of the disorder are often described in subjective assessments of the disorder, but not always easily quantifiable using objective tests. This paper investigates a speech characteristic not yet assessed in FTLD patients, that of "temporal cycles" [Henderson *et al.* (1966); Butterworth and Goldman-Eisler (1979)]. Temporal cycles in the speech of healthy adults consist of alternating and roughly equal periods of fluent and hesitant speech [Roberts and Kirsner (2000)]. A time series analysis of temporal cycles was conducted using spontaneous speech from 45 adults diagnosed with FTLD. Patients' cognitive functioning was assessed using the language-specific clinical dementia rating (CDR) scale by board-certified neurologists. Periodicity of temporal cycles was quantified using the proportion of the energy in the highest peak to the total energy in the power spectrum. It was found that this measure was correlated with independent CDR assessments. The results of this study indicate that temporal cycles may be used to characterize the effects of neurodegenerative disorders on speech communication. [Work supported by US NIA Grants Nos. R01-AG023195, P50-AG16574, P30-AG19610, and NIH NIA-1R01AG026390, and Univ. of MN Academic Health Center.]

4pSCb2. Transition characteristics in speakers with dysarthria and in healthy controls: Part IV: Additional data on vital capacity transitions and stroke patients. Gary Weismer, Christina Kuo, and Phoebe Allen (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706, gweismer@wisc.edu)

Formant transitions are known to provide important cues for speech perception, sound identification, and inferences to articulatory behavior. This study describes and examines three types of formant transitions [consonant-vowel (CV), vowel-consonant (VC), and diphthong transitions] in four groups of speakers: healthy, ALS, Parkinson's disease, and stroke. This is an extension from previous work by Weismer *et al.* [J. Acoust. Soc. Am. **121**, 3135 (2007)] and Weismer *et al.* [J. Acoust. Soc. Am. **124**, 2558 (2008)], who showed shallower slopes for CV and diphthong transitions in persons with dysarthria (ALS and PD). To better understand the characteristics of the different transition types in healthy and disordered populations, two questions are addressed here. First, are CV transitions associated with dysarthria different from those in healthy speakers in a way comparable to the observed differences in diphthong transitions? Second, do VC transitions show the same normal characteristics, and are the differences between healthy speakers and speakers with dysarthria the same as in CV transitions? Distributional analyses for 50-ms CV, 50-ms VC, and diphthong (highest-lowest/lowest-highest) F2 transition measures will be presented for 18 healthy speakers, 4 speakers with ALS, 4 speakers with PD, and 20 stroke patients. [Work supported by NIDCD R01 DC003723.]

4pSCb3. Measuring speech and language characteristics of effects of medications on cognition. Serguei Pakhomov, Susan Marino, Angela Birnbaum, Chamika Hawkins-Taylor, and Ilo Leppik (Serguei Pakhomov, Univ. of Minnesota, Minneapolis, MN 55455, pakh0002@umn.edu)

Many of the drugs prescribed for various diseases and syndromes, including chronic pain and epilepsy, significantly impair cognitive functioning with large individual variation. Language is a highly individualized and directly observable product of human cognition that is central to our everyday functioning. A system for automated language and speech analysis (SALSA) was developed that relies on using an automatic speech recognition engine (HTK 3.4) [Young *et al.* (2008)] to perform forced-alignment between the audio of spontaneous speech samples and their transcripts to measure a number of speech and language characteristics including fluency, speaking rate, change in fundametal frequency, and information content of spontaneous narratives. The system was piloted on a population of 14 healthy volunteers who participated in a randomized, placebo-controlled study of cognitive effects of an anti-epileptic medication (topiramate). Our preliminary results suggest that SALSA captures a number of fluency and lexical characteristics sensitive to the effects of topiramate, thus providing an objective mechanism to quantify the degree of cognitive impairment in individulas affected by medications. Our current results are consistent with prior work investigating speech and language correlates of mild cognitive impairment [Roark *et al.* (2007)] and fronto-temporal dementia [Pakhomov *et al.* (2009)]. [Work supported by a grant from the Univ. of Minnesota Academic Health Crt.]

4pSCb4. The differential effect of altered auditory feedback on speech in early- and late-onset of Parkinson disease. Emily Q. Wang (Dept. of Commun. Disord. and Sci., Rush Univ. Medical Ctr., 1653 West Congress Parkway, 203 SENN, Chicago, IL) and Leo Verhagen Metman (Rush Univ. Medical Ctr., Chicago, IL)

Advanced speech symptoms in Parkinson disease (PD), such as "festinating speech" or palilalia and frequent hesitations, are a clinical challenge. This report is a part of a larger study testing the hypothesis that the use of altered auditory feedback (AAF) would improve speech intelligibility in patients with PD and more advanced speech impairment. Our initial report showed that the use of AAF indeed improved speech intelligibility in these patient's spontaneous speech [Wang *et al.* (2008)]. In this report, the spontaneous speech samples produced by four older-onset (aged 72.4 SD 6.02 years) vs four young-onset PD patients (aged 54.2 SD 4.44 years), with and without the use of AAF, were analyzed. The results indicated that, regardless of the use of AAF, the older-onset patients performed poorer than the young-onset patients in many areas including the mean length of utterance in words, number of different word roots, within utterance pause time, and words per minute. These findings are consistent with the report that patients with an older age at onset had more rapid progression of PD than those with a younger age at onset in mentation, freezing, and activities of daily living. Possible underlying mechanism for the findings will be discussed.

4pSCb5. Temporal analysis of simultaneous nasopharyngoscopic and nasometric recordings. Wei Tian (Dept. of Hearing and Speech Sci., Univ. of Maryland, 0141D Lefrak Hall, College Park, MD 20742)

The velopharyngeal port closes and opens by the velar and pharyngeal wall motion during speech. Velopharyngeal dysfunction (failure to close or open appropriately) often involves insufficient motion or mistiming of the motion. Despite the fact that evaluation of the velopharyngeal port size by nasopharyngoscopy has been popular in clinical settings, the temporal pattern of velar and pharyngeal motion has never been studied. On the other hand, separate nasal and oral recording of speech acoustics with a Nasometer is reported to demonstrate speech nasality change in customized utterances. The present study aimed at investigating the correlations between velopharyngeal motion and speech acoustic change by simultaneous recordings of nasopharyngoscopy and nasometry. Kymographic analysis of the motion in the velum and lateral pharyngeal walls were performed in two female adult speakers during five repetitions of six VCVNVCV at a controlled speech rate. Although the velar and pharyngeal motions show the similar pattern, they do not initiate and end simultaneously. There is strong correlation between the physiologic motion and acoustic change, but they are not synchronized and have very different temporal patterns. There is also difference across speakers and phonemes.

THURSDAY AFTERNOON, 22 APRIL 2010 ESSEX A/B/C, 1:30 TO 2:55 P.M.

Session 4pSP

Signal Processing in Acoustics, Underwater Acoustics, and Architectural Acoustics: Maximum Entropy and Bayesian Signal Processing II

Zoi-Heleni Michalopoulou, Cochair
New Jersey Inst. of Technology, Dept. of Mathematics, Newark, NJ 07102–1982

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., Architecture, 110 8th St., Troy, NY 12180

Invited Papers

1:30

4pSP1. Bayesian parameter estimation in adaptive psychometric procedures: Simulated and experimental results. Jeremiah J. Remus (Dept. of Elec. and Comput. Eng., Clarkson Univ., Box 5720, Potsdam, NY, 13699) and Leslie M. Collins (Duke Univ., Durham, NC 27708)

The wide use of psychometric assessments and the associated time required to conduct such experiments have motivated a substantial research effort focused on developing more efficient psychometric procedures to expedite the estimation of parameters of interest. Adaptive step size psychometric procedures, which have increasingly received attention as an alternative to fixed step size procedures, can be viewed as having two separate stages: estimation of the parameters of interest based on the available data and a selection of the experimental settings in order to maximize the information obtained from the next trial. Psychometric procedures that utilize a Bayesian framework for parameter estimation have become more prevalent in recent years. In this talk, we will present experimental results for Bayesian psychometric procedures from both psychoacoustics studies and computer simulations. Several com-

puter simulations explored the sensitivity of the procedure to different model parametrizations and evaluated performance using stimulus selection rules based on minimizing entropy and maximizing information gain. Psychoacoustic studies using two different listening tasks were used to investigate the robustness and stability of estimates provided by the Bayesian parameter estimation procedure. Collectively, the outcomes suggest that a Bayesian framework may be an important component in the development of more efficient psychometric procedures.

1:50

4pSP2. Nested sampling for room-acoustics energy decay analysis. Ning Xiang, Tomislav Jasa, and Jonas Braasch (Graduate Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY)

Room-acoustic energy decays often exhibit single-rate or multiple-rate behavior in a wide variety of enclosures. There has been an ever increasing need to investigate where and under what conditions the single-rate and or multiple-rate energy decays occur and to quantify/characterize the energy decay process by estimating a set of decay parameters. In this architectural acoustics application, both energy decay model selection and decay parameter estimation are of practical significance. This paper discusses a model-based sound energy decay analysis within a Bayesian framework and applies the most recent Bayesian sampling method (nested sampling) to sound energy decay analysis. The nested sampling reverses the historical Bayesian computation approaches, yielding the Bayesian evidence as the prime target with representative posterior samples available as optional by-products. Bayesian evidence is crucial for model selection, while the representative posterior samples are of central importance for parameter estimation. Taking the energy decay analysis in architectural acoustics as an example, this paper demonstrates that two different levels of inference, decay model-selection and decay parameter estimation, can be cohesively accomplished by nested sampling.

Contributed Papers

2:10

4pSP3. Comparison of fading statistics for shallow and deep acoustic sources in a continental shelf environment. Alexander W. Sell, R. Lee Culver, Colin W. Jemmott, and Brett E. Bissinger (Graduate Program in Acoust., Penn State Univ., 201 Appl. Sci. Bldg., University Park, PA 16802, aws164@psu.edu)

Previous work has shown a noticeable difference in the effects of the ocean environment on signals from shallow and deep moving sources. These effects are seen in received amplitude statistics, and such statistics can be used in passive acoustic depth classification. This talk presents a statistical analysis of signals from a September 2007 shallow water acoustic transmission test performed along the continental shelf off the coast of southeast Florida. The data used include low frequency (between 25 and 450 Hz), continuous-wave signals from a towed source at 100 m depth, as well as tones from surface ships in the area. Using statistical class models in a Minimum Hellinger Distance Classifier, the usefulness of received signal amplitude statistics for passive acoustic source-depth classification is discussed. [Work supported by ONR Undersea Signal Processing.]

2:25

4pSP4. Depth discrimination in shallow water by matched-field tracking. Donald R. DelBalzo and James H. Leclere (QinetiQ North America, Hwy. 190 East, Slidell, LA 70461, donald.delbalzo@qinetiq-na.com)

Passive acoustic sonars have difficulty in detecting quiet sources in noisy shallow-water environments. The standard approach to improving detection performance is to use a large aperture linear array of horizontally distributed hydrophones and azimuthal processing to increase the signal-to-noise ratio (SNR) through beamforming. A newer approach is to apply matched-field correlation techniques and estimate source locations in depth and range using signals received on a linear vertical array. These localization techniques are not designed for detection of quiet targets and some potential applications preclude the use of vertical arrays for practical reasons, like damage caused by fish trawling. In this work, we examine the utility of matched-field processing on linear and planar horizontal arrays on the ocean bottom in shallow water. We employ a track-before-detect strategy to convert matched-field localizations into submerged source detections in various physical and acoustic environments. We show the advantages of planar compared to linear arrays to discriminate submerged from surface sources as a function of SNR.

2:40

4pSP5. A feasibility study on the low probability of intercept sonar. J. Daniel Park, David J. Miller, John F. Doherty (Dept. of Elec. Eng., The Penn State Univ., University Park, PA 16802), and Stephen C. Thompson (The Penn State Univ., State College, PA 16804)

The feasibility of low probability of intercept for sonar is explored. Using a noise-like active sonar signal, the transmitter (platform) employs a matched filter for echo detection while the target is assumed to use an energy detector. Decision statistic distributions are developed and detection performances are compared with a previous work (Gaussian distribution assumed). We then explore the detection advantage the platform can achieve by evasive on-off keying and by optimization of transmitted power. A favorable operating region of the platform in the (low power, small range) region of the (range, power) plane is identified. This suggests that the platform should start detection using a low-power probing signal, increasing power until a reliable detection rate is first achieved while ensuring that the detection rate at the target does not exceed a specified level. Second part of the feasibility study is about covert range estimation. Two matched-filter based algorithms were developed for the platform to obtain credible range estimations while ensuring the target fails to detect the platform. A platform-target encounter scenario was designed for detailed waveform-based Monte Carlo simulation. Characterization of variables in the model was performed to show the feasible conditions of covert range estimation. [Work sponsored by ONR, ULI.]

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

P. Battenberg, Chair, ASC S1
Quest Technologies, Inc., 1060 Corporate Center Dr., Oconomowoc, WI 53066 4828

R. J. Peppin, Vice Chair, ASC S1
Scantek, Inc., 6450 Dobbin Rd., #A, Columbia MD 21045

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43, Acoustics and IEC/TC 29 Electroacoustics, take note - those meetings will be held in conjunction with the Standards Plenary meeting at 8:00 a.m. on Thursday, 22 April 2010.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

C. A. Champlin, Chair, ASC S3
University of Texas, Dept. of Communication Sciences & Disorders, CMA 2-200, Austin, TX 78712

D. A. Preves, Vice Chair, ASC S3
Starkey Laboratories, Inc., 6600 Washington Ave., S., Eden Prairie, MN 55344

Accredited Standards Committee S3 on Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43, Acoustics and IEC/TC 29 Electroacoustics, take note - those meetings will be held in conjunction with the Standards Plenary meeting at 8:00 a.m. on Thursday, 22 April 2010.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3/SC 1, Animal Bioacoustics

D. K. Delaney, Chair, ASC S3/SC 1
USA CERL, 2902 Farber Dr., Champaign, IL 61822

M. C. Hastings, Vice Chair, ASC S3/SC 1
6422 Crosswoods Dr., Falls Church VA 22044 1214

Accredited Standards Committee S3/SC 1 on Animal Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

Scope of S3/SC1: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

THURSDAY EVENING, 22 APRIL 2010 7:30 P.M.

OPEN MEETINGS OF ASA TECHNICAL COMMITTEES

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held immediately after the Social Hours. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings, including proposals for special sessions, workshops, and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

Committees meeting on Thursday are as follows:

Animal Bioacoustics	Grand Ballroom I/II
Architectural Acoustics	Grand Ballroom III/IV
Biomedical Ultrasound/Bioresponse to Vibration	Kent A/B
Speech Communication	Laurel A/B
Underwater Acoustics	Laurel C/D

Session 5aAA

Architectural Acoustics, Noise, and INCE: Case Studies, Applications, and Integration of Architectural Acoustics in Building Information Modeling Three-Dimensional Modeling

Norman H. Philipp, Chair
Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0791

Chair's Introduction—7:30

Invited Papers

7:35

5aAA1. Building information management: An acoustics consultant's perspective. Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Architecture and engineering firms use building information management (BIM) software applications increasingly in the design process; similarly, building owners create BIM models for existing facilities both for internal use and to aid future tenants, facilities managers, and designers in planning for maintenance programs, restoration, or renovation. As for other design and engineering disciplines, these tools provide powerful opportunities to the acoustics community. The primary benefits to architectural acoustics consultants fall into two broad categories: first, in the construction process by providing a means of organizing the specifications for materials and constructions, and second, in the design process, by establishing a framework of data and acoustical attributes that in the future could potentially aid in acoustical calculation and analysis.

7:55

5aAA2. Building information modeling with revit architecture and acoustical coordination. Katherine LePage (GWWO, Inc. Architects, 800 Wyman Park Dr., Ste. 300, Baltimore, MD 21211, klepage@gwwoinc.com)

The adoption of BIM has provided a huge shift within the architectural design community. It provides an opportunity for more thoughtfully designed buildings as well as comprehensive coordination with consultants. This presentation will cover the basics of how programs like revit architecture are being used by the architectural community in general and more specifically for acoustical coordination. Topics covered will include visualization, basic modeling, consultant coordination, and construction administration. The work of GWWO Inc./Architects will be used as a basis for providing real world examples of how the program has been used through all stages of design and construction. The 70 000-sft^2 Daniel Z. Gibson Performing Arts Center at Washington University will be a featured project. The center provides a home for both the Music and Drama Departments and includes state of the art teaching and performance spaces. Challenges for the project included transforming an existing 600 seat house theater into a more intimate 440 seat theater with a balcony level, designing an acoustically pleasing 200 seat recital hall while minimizing the sound focusing of a concave glass wall along one side of the hall and the addition of a 175 seat flexible experimental theater.

8:15

5aAA3. Revit and the role of architectural acoustics in the design process. Jeff Larrick (Autodesk, West Palm Beach, FL 33401, jeff.larrick@autodesk.com)

A presentation by Autodesk discussing the role of Revit in the building information modeling design process and how it allows for collaboration between the architect and acoustical consultant.

Contributed Papers

8:35

5aAA4. Distilling the acoustical model from BIM (Building Information Modeling) Standard architectural, mechanical and structural models: Robust acoustical templates, limitations, and recommendations. Richard A. Vedvik and Jon W. Mooney (KJWW Eng. Consultants, 623 26th Ave., Rock Island, IL 61201, vedvikra@kjww.com)

The integration of architectural acoustics into building integration modeling (BIM) software and the implementation of such integration is discussed between available systems and platforms. The process of integrating the various platforms of building integration modeling software, in such a manner that the assigned properties can be shared, manipulated, and exported, is of great interest to acousticians. As the architectural model is created, a standardization of materials and generic construction methods is desirable so the relevant properties assigned to the various building components are able to be imported and adjusted by robust acoustical templates (RATs). Additional complications are identified as the acoustical model is detached from the live model. Standardization is also important between the architectural, structural, and mechanical models in order to create a common shared model that is able to maintain the relevant properties, both physical and analytical. Finally, the operation of a RAT prototype is demonstrated in the distillation of an acoustical model from standard BIM architectural, mechanical, and structural models.

8:50

5aAA5. Error estimation due to sample size effects of *in situ* surface impedance measurements. Eric Brandao, Erico Fulco, Arcanjo Lenzi (Federal Univ. of Santa Catarina, Joao Pio Duarte Silva St., 250, Bl3, Ap 201 Corrego Grande Florianopolis-SC, 88037-001 Brazil, eric@lva.ufsc.br), and Emiel Tijs (Microflown Technologies)

The boundary element method (BEM) is used to predict the effect of the sample size on *in situ* impedance measurements on a locally reactive sample

5a FRI. AM

at normal incidence. The BEM model is based on a direct measurement of the acoustic impedance above the sample (with a combined sound pressure and particle velocity sensor). As the sound source is placed close to the sensor, the acoustic waves cannot be considered plane anymore. Because of that, two models to obtain the surface impedance based on the acoustic field are presented and compared. It is also shown that the sample size, the thickness of the sample, and the positions of the source and receiver affect the measured surface impedance and the calculated absorption coefficient. Those effects are parametrized in a way that the user of the *in situ* measurement setup may have an estimate of the required sample size for a desired frequency range with a given error allowance.

9:05

5aAA6. Examining the relationship between total acoustic absorption and late lateral energy (GLL). David A. Dick and Michelle C. Vigeant (Dept. Mech. Eng. Acoust. Prog. and Lab., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

Listener envelopment, the sense of being immersed in the sound field, has been shown to correlate with the objective parameter late lateral energy (GLL). Barron proposed that GLL is primarily a function of the ratio of reverberation time to auditorium volume or the total acoustic absorption [Barron, Appl. Acoust. **62**, 185–202 (2001)]. An investigation has been conducted to examine this proposed theory with measurement data and predicted values from a room acoustics computer modeling program. Room impulse response measurements have been taken in the Belding Theatre in downtown Hartford, CT. Features of the hall include variable absorption in the form of curtains, both within the hall and also within small adjacent coupled volumes. Measurements were taken in several absorption configurations. A detailed room acoustics computer model of the hall has been created in ODEON v9.22 to examine the accuracy of the model's prediction of GLL. The model was initially validated using the measured parameters of reverberation time (T30), early decay time (EDT), and clarity index (C80). A comparison of the measured and predicted results will be discussed, along with an analysis of the relationship between GLL and total acoustic absorption. [Work supported by a University of Hartford Greenberg Junior Faculty Grant.]

9:20

5aAA7. Classroom acoustics in the developing world: A student project to develop simple assessments and treatments. Nathan McNeill, Matthew Blevins, Emily Drott, Alex Kremer, Mark Kusch, Brian Pluhar, Ashley Ringer, Valerie Sargent, and J. Stuart Bolton (Purdue Univ., Neil Armstrong Hall of Eng., Rm. 1300, 701 W. Stadium Ave., West Lafayette, IN 47907-2016, nmcneill@purdue.edu)

An inadequate educational system can be crippling to a developing country struggling to advance in the modern world. Poor classroom acoustics are a significant factor that can contribute to an inadequate education. Classroom acoustics is a mature field of research in the developed world; however, the classroom conditions and resources found in the developing world present new research challenges. A special project course was developed at Purdue University to give undergraduate students the opportunity to learn about and address classroom acoustics problems in underdeveloped regions. The goal of this special project course was to develop an evaluation and design manual that can be distributed to teachers and educational officials in those regions. The purpose of the manual is to help identify acoustic problems in classrooms, provide suggestions of simple noise control solutions, and outline best practices for the design of new classrooms. The manual specifically addresses problems related to excessive reverberation, poor signal-to-noise ratios, and noise transmission from outside the classroom. The ultimate intent of this project was to increase awareness of the unique classroom acoustic conditions in the developing world.

9:35

5aAA8. Experimental validations of the transport equation model for room-acoustic predictions in long spaces. Yun Jing and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

This paper presents an experimental study on steady-state sound pressure levels and reverberation times in a tenth long-space scale model. The purpose of this work is to validate a recently developed one-dimensional transport equation model for acoustic predictions in long spaces. A series of measurements has been conducted in the scale model to reveal receiver-position dependence of measured and modeled parameters in the long spaces. Experimental results at different octave bands are obtained. This paper will compare the acoustically measured results with that of simulations.

9:50

5aAA9. An experimental scaled-model for coupled volumes with an automated high-spatial-resolution scanning system. Joonhee Lee, Craig Schaefer, and Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, New York 12180, gi20plus@gmail.com)

This paper presents an experimental study of low-frequency behavior in coupled volumes. In order to examine low-frequency behavior in coupled volumes, an eighth-scale model of two coupled volumes with an automated high-spatial-resolution scanning system has been developed. The validated scanning system makes it possible to acquire thousands of room impulse responses on a designated planar grid with a fine grid size in the scaled model. The high-spatial-resolution scanning results are used to investigate the characteristics of acoustic wave propagation in both single-space and coupled-volume systems. This paper discusses some design issues relevant to high-quality, high-spatial resolution scanning, to diffusely reflecting interior surfaces, and to variable coupling aperture sizes. Some representative scanning results of sound pressure propagations and sound energy flows will also be demonstrated.

10:05—10:20 Break

10:20

5aAA10. Some preliminary investigations on the use of a diffusion equation model for room-acoustic parameters prediction. José Escolano (Dept. Telecommunication Eng., Univ. of Jaén, 23700, Linares, Spain, escolano@ujaen.es), Juan M. Navarro (San Antonio's Catholic Univ. of Murcia, 30107 Guadalupe, Spain), José J. López (Universidad Politécnica de Valencia, 46021 Valencia, Spain), and Ning Xiang (Rensselaer Polytechnic Inst., Troy, NY 12180)

Over the very recent years, an alternative model to describe sound fields in enclosures, based on the mathematical theory of diffusion, has been proposed. This model is based on the concept of sound particles traveling along straight lines and striking walls or scattering objects existing in a room. The scope of this work is to present some preliminary investigations in the prediction of objective room-acoustics parameters, such as reverberation time, clarity, and definition through the use of a diffusion equation. This paper will discuss modeling accuracy of the room-acoustics parameters using some numerical examples. Moreover, some issues concerning the numerical implementation by means of finite difference schemes for the diffusion equation are addressed.

10:35

5aAA11. Micro-perforated stretched foils as sound absorbers and barriers. Christian Nocke, Catja Hilge (Akustikbuero Oldenburg, Katharinenstr. 10, 26121 Oldenburg, Germany, nocke@akustikbuero-oldenburg.de), and Jean-Marc Scherrer (BARRISOL S.A.S, F-68680 Kembs, France)

A typical set-up of a micro-perforated sound absorber is a micro-perforated sheet or panel in front of a closed air cavity. The sound absorption coefficient of these set-ups can be easily calculated with a high accuracy according to the well-known approximation of D.-Y. Maa if all defining geometrical parameters (diameter of microperforation, distance between orifices, panel thickness, and air cavity depth) are known. Other applications are freely suspended sails of micro-perforated sheets as well as double layer structures mad of micro-perforated materials. In this contribution measured sound absorption coefficients of different set-ups as well as measurements of level decrease will be presented and discussed. For some of these assemblies

no closed calculation model exists so far. Comparisons between computer-based calculations of layered absorber set-ups show good results for some assemblies. Finally room acoustic applications of micro-perforated stretched sheets are shown.

10:50

5aAA12. Insertion loss estimation of an opened enclosure using a statistical ray-tracing model. Buye Xu, Scott D. Sommerfeldt, and Timothy W. Leishman (Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84606, buye.xu@gmail.com)

Estimating the insertion loss (IL) of an enclosure with one or more openings is important for many noise control applications. In the low-frequency range (below the Schroeder frequency) where modal effects are dominant, numerical simulations based on modal expansions or the finite element method are feasible for obtaining the desired predictions. Above the Schroeder frequency, statistical room acoustics theories are more desirable because numerical simulations are not applicable or are computationally expensive. However, existing room acoustics theories can only deal with simple cases where a diffuse field can be assumed. The acoustic field inside enclosures is generally sufficiently complicated in practice such that the diffuse field assumption is not valid. Examples of the complex conditions include narrow spatial regions, non-uniform boundaries, and the existence of obstacles in the enclosure. In this paper, a model that adopts the basic ideas of both ray-tracing and Sabine's room acoustics theory will be introduced to deal with the difficulties mentioned. This new model greatly improves the ability to estimate acoustic performance when the enclosure conditions are not ideal, yet requires much less computation than the normal ray-tracing model. Comparisons of different models with experimental data will be presented.

11:05

5aAA13. High-resolution absorption mapping with a pu surface impedance method. Emiel Tijs (Microflown Technologies, 6826 CC Arnhem, Arnhem, The Netherlands, tijs@microflown.com)

The *in situ* method to measure the surface impedance with a pu probe documented is published in many publications. The method is based on the measurement of sound pressure (p) and particle velocity (u) close to an acoustic absorbing material. A loudspeaker at a defined distance is applied to generate a sound field with a known radiation impedance. The impedance of a small area (a few square cm^2) with a known impedance is scanned with an ultraminiature pu probe very close to the surface. The area is made of steel with a cut-out, behind this a material with a known impedance is placed. In this paper the method is explained, the spatial accuracy of the measurement is examined, and a visualization technique is presented with a display of the spatial distribution (two dimensional picture) of the damping properties as function of frequency. Finally the high-resolution measurements are combined and compared with a method that covers a larger surface area.

11:20

5aAA14. Is it 6 decibels? The effect of pressure doubling on exterior to interior transmission loss values. Noel Hart, Kimberly Lefkowitz, Arno S. Bommer, and Robert D. Bruce (CSTI Acoust., P.O. Box 218808, Houston, TX 77218-8808, noel@cstiacoustics.com)

When calculating sound levels inside a building due to external sources, should some value be added to the expected sound level? ATSM E966 recommends adding 6 dB to this level. It has been theorized that this is due to pressure doubling. The effect of pressure doubling on the calculated exterior to interior sound transmission in the direct field is, however, often misunderstood. Most texts on the subject simply list a 5- or 6-dB correction factor that is applied to the transmission loss value. The reasoning of this particular number is neither expounded upon nor explained. This paper investigates and explains this correction factor.

11:35—12:00 Panel Discussion

FRIDAY MORNING, 23 APRIL 2010 GRAND BALLROOM I/II, 8:45 A.M. TO 12:00 NOON

Session 5aAB

Animal Bioacoustics, Signal Processing in Acoustics, and Noise: Topical Meeting on Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication III: Incorporating Contextual Features into Automated Detection and Classification Algorithms

Ann E. Bowles, Cochair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Sean K. Lehman, Cochair
Lawrence Livermore National Lab., Livermore, CA 94550

Chair's Introduction—8:45

Invited Papers

8:50

5aAB1. Multipath features as a contextual clue for classification. Christopher O. Tiemann (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78731)

A single odontocete click is often heard as multiple click echoes due to the many paths sound travels between the source (whale) and receiver; these echoes are called multipath arrivals, and together they form an arrival pattern. Because each arrival can appear differently due to propagation effects, efforts to associate recorded clicks with individual whales can be frustrated by the presence of multipath arrivals, particularly when many whales click simultaneously. Automated classifiers may associate arrivals from one click event into many groups, even though they come from one source, thus overestimating the number of whales in a recording. However, collectively a multipath arrival pattern can be exploited to huge advantage when trying to do individual association. Typically, the relative spacing

of arrivals changes slowly over time (due to slow movement of the source) and is unique for a given source location. The persistence of stable arrival patterns over successive click events provides a contextual clue for associating many arrivals to a single individual or can validate automated classification techniques. Finding unique arrival patterns should provide an accurate estimation of the number of sources in an area and even serves as the first step in passive acoustic tracking of whales.

9:10

5aAB2. Contextual rhythmic analysis of beaked whale clicks for passive acoustic identification. Natalia Sidorovskaia, Philip Schexnayder (Dept. of Phys., Univ. of Louisiana at Lafayette, UL Box 44210, Lafayette, LA 70504-4210, nas@louisiana.edu), George E. Ioup, Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148), Christopher O. Tiemann (Univ. of Texas at Austin, Austin, TX 78713), Alexander E. Ekimov, and James Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Oxford, MS 38677)

The Littoral Acoustic Demonstration Center (LADC) has collected an extensive acoustic library of marine mammal phonations in the Northern Gulf of Mexico, including the Cuvier's and Blainville's beaked whale clicks. An automatic algorithm, which performs cadence frequency analysis, has been developed and successfully applied to sperm whale clicks to identify temporal evolution of individual rhythmic patterns in overlapping trains of simultaneously clicking whales in a group. In this study the results of the association of cadence frequencies with individual beaked whales in a group for the LADC 2007 Gulf of Mexico recordings are reported. The algorithm is a part of an integrated tool proposed by LADC for identification of individual animals in a group. The cadence frequency association results are compared to multi-attribute self-organizing map clustering and to time-of-arrival analysis and localization. A multi-band version of the algorithm is also presented. It provides species association cues for complex cluttered acoustic environments where different species phonate with similar inter-click intervals. The proposed approach is beneficial for passive acoustic studies of beaked whale populations and social behavior and may contribute to understanding acoustic communication of these evasive animals. [Research supported by ONR and SPAWAR.]

9:30

5aAB3. Using click change detection to modify cluster analysis for acoustically identifying individual sperm whales with changing aspect. George E. Ioup, Juliette W. Ioup, Lisa A. Pflug (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, geioup@uno.edu), Christopher O. Tiemann (Univ. of Texas at Austin, Austin, TX), and Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA)

As described by Mohl *et al.* J. Acoust. Soc. Am. **114**, 1143–1154 (2003)], the acoustic properties of the received clicks of a turning sperm whale change as the whale rotates. These changing properties (the time signal, the discrete Fourier transform, and the discrete wavelet transform are studied here) confound cluster analysis attempts to identify individual whales by grouping like clicks in classes representing individuals. The 4th International Workshop on Detection, Classification and Localization of Marine Mammals Using Passive Acoustics has associated workshop data containing a sequence of sperm whale clicks which most likely come from the same individual (as shown by careful manual click analysis) turning with respect to the acoustic recording buoy. The time domain sequence of clicks indicates that the angle of a line from the head of the whale to the buoy appears to be changing from greater than 90 deg to less than 90 deg and back again. Time evolution and click change detection are applied to follow the whale and give correct automatic identification. This analysis is incorporated into the clustering process. [Research supported by ONR and SPAWAR.]

9:50

5aAB4. Acoustic identification of individual beaked whales using self-organizing maps for automated classification of clicks in Littoral Acoustic Demonstration Center (LADC) data. Juliette W. Ioup, George E. Ioup, Lisa A. Pflug (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, jioup@uno.edu), Natalia A. Sidorovskaia, Philip Schexnayder (Univ. of Louisiana at Lafayette, Lafayette, LA), Christopher O. Tiemann, and Alan Bernstein (Univ. of Texas at Austin, Austin, TX)

The Littoral Acoustic Demonstration Center (LADC) acquired underwater acoustic data from clicking sperm and beaked whales in the Gulf of Mexico during an experiment in July 2007. Analysis of sperm whale echolocation and coda clicks in the data has been presented at previous ASA meetings. Research emphasis is now on identifying individual beaked whales from their echolocation clicks. Again, clustering methods present one appropriate approach to the automated identification of these whales by computer. Self-organizing maps have the ability to limit the number of clusters automatically, and their successful use with sperm whales encourages the application to beaked whales as well. Results for time, frequency, and wavelet domain data allow comparison among clustering done using different signal representations and help highlight areas of concern in the analysis. As in the case of sperm whales, the beaked whale identifications are verified by comparing to cadence analysis (as presented in an accompanying abstract) and to a limited extent by manual analysis and localization. [Research supported by ONR and SPAWAR.]

10:10—10:25 Break

Contributed Paper

10:25

5aAB5. Automated detection of sperm whales (*Physeter macrocephalus*) creaks and association with depredation events. Delphine Mathias, Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr. La Jolla, CA, 92037-0238, delphine.mathias@gmail.com), Jan Straley (Univ. of Alaska Southeast, Sitka, AK 99835), John Calambokidis, Gregory S. Schorr (Cascadia Res. Collective, Olympia, WA, 98501), and Bill Burgess (Greeneridge Sci. Inc., Goleta, CA 93117)

Sperm whales have been depredating black cod (*Anoplopoma fimbria*) from demersal longlines in the Gulf of Alaska for decades, but the behavior has now become pervasive enough that it is starting to affect government estimates of the sustainable catch, motivating further studies of this behavior. In July 2007 and June 2009, bioacoustic tags were attached to adult sperm whales off Southeast Alaska under both natural foraging conditions and situations wherein the animals were depredating sablefish from commercial longlining vessels. Tags remained attached for a cumulative hours, yielding more than 100 h of depth, orientation, and acoustic data. Results show that when depredating, sperm whales often dive at depths shallower than 50 m, compared to natural foraging depths of 300–400 m in the area. Dive durations, pitch, and roll rates are also distinctively different. Techniques such as spectrogram cross-correlation and cepstral analysis are used to automatically detect creak events, which have lower intensities and narrower frequency ranges than conventional clicks. Acoustic measurements such as creak production rate, duration, and intensity are then extracted. Correlations are derived between creak events, sudden changes in body orientation and depredation rate, with a goal of determining an acoustic measure of depredation activity and efficacy. [Work conducted under the SEA SWAP Program, supported by the North Pacific Research Board and the National Geographic Society.]

Invited Paper

10:40

5aAB6. Detecting sequences of calls. David K. Mellinger (Coop. Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, david.mellinger@oregonstate.edu)

Many sound-producing animals make vocalizations with predictable timing patterns. In some cases, animals produce long, regular sequences of sounds, either for breeding (e.g., many baleen whales) or for foraging (e.g., most toothed whales and dolphins); in other cases, different call types are produced in succession with the intervals between calls following certain timing patterns. Here I describe methods for detecting these timing patterns. Both of these methods operate on a "detection function," the estimated likelihood at each instant of time that a call is present. Long, regular sequences can be detected using an autocorrelation of the detection function that significantly increases the detectability of sequences compared to single calls. Different call types occurring in irregular but predictable timed patterns can be detected using methods similar to dynamic time warping; again, the detectability of the patterns is notably higher than the detectability of single calls. [Work funded by ONR and the Navy's Environmental Readiness Division.]

11:00—12:00 Panel Discussion

FRIDAY MORNING, 23 APRIL 2010 LAUREL C/D, 8:25 A.M. TO 12:00 NOON

Session 5aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasonic Characterization of Bone I

Keith A. Wear, Chair
Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993

Chair's Introduction—8:25

Invited Papers

8:30

5aBB1. Extracting bone strength information from x-ray image data and applications in clinical research. Thomas J. Beck (Dept. of Radiology, Div. Medical Imaging Phys., Johns Hopkins Univ. School of Med, Baltimore, MD 21287)

Osteoporosis is a symptomless disease that reduces bone strength so that bones fracture easily, but it is often undiagnosed until fractures occur. Because effective treatments exist, it is important to detect the disease before fractures occur. Current diagnostic methods measure some observable parameter that is statistically correlated with the presence of disease. Most commonly the parameter is a form of bone density although ultrasonic parameters have also been used. Engineers have designed some excellent instruments that measure parameters with high accuracy and precision but even the best methods identify less than half of those at risk of fractures. This is not surprising when discrimination is based how average patients differ from normal rather than on what is actually wrong with them, i.e., reduced bone strength. Mechanical engineering methods can be used to predict bone strength if the material strength and the structural geometry can be measured. Currently there are no reliable non-invasive methods for measuring bone material properties so current mostly x-ray based methods measure structural geometry alone. A review of current methods will be presented together with an overview of what they tell us about osteoporosis as well as what we still need to know.

8:45

5aBB2. Quantitative ultrasound in the management of osteoporosis. Marc-Antoine Krieg and Hans Didier (Ctr. of Bone Diseases, CHUV, Av. Pierre-Decker 4, 1011 Lausanne, Switzerland)

Bone mineral density (BMD) is one of the major determinants of bone strength and fracture risk (FR), but it is not the only determinant, and overlap exists in BMD values between patients with and without fractures. It is therefore important to improve the strategy in better detecting patients at high FR who will benefit from treatments (TT). Quantitative bone US(QUS) are portable and less expensive than DXA. Only few QUS measuring the heel have reached a sufficient level of validation to be routinely used. They usually measure the attenuation (BUA) and the velocity (SOS) of the US wave. It is well established that valid heel QUSs predict fragility fracture in PM women and men >65, independent of BMD. Then, heel QUS can be used in combination with other clinical risk factors (CRFs) in a model of risk assessment. Then, when DXA is available, the primary clinical use of heel QUS is to identify, in conjunction with CRF, a population at low FR in which no further diagnostic evaluation is necessary. On the other hand, if central DXA cannot be done, pharmacologic TT can be initiated if the probability, as assessed by heel QUS in conjunction with CRF, is sufficiently high.

9:00

5aBB3. New developments in x-ray micro-computed tomography technology and its applications to study bone micro-architecture. Rasesh Kapadia (Scanco USA, Inc., 985 Old Eagle School Rd., Ste. 511, Wayne, PA 19087)

Developments in hardware and software capabilities, along with a significant increase in access to equipment and awareness of the methods, over the last 10 years have enabled micro-computed tomography (CT) technology to establish itself as one of the basic tools to study bone micro-architecture in the laboratory. A wide range of equipment is now available ranging from synchrotron based facilities to compact desktop scanners for small to large specimens and scanners for *in vivo* pre-clinical imaging. The various scanners offer the ability to acquire images with resolution ranging from sub-micron to 100 μm in acceptable scan times and where applicable, with acceptable x-ray doses. More recently, high-resolution scanners (under 100 μm) have become available for clinical research. Micro-CT is now widely used to investigate bone from pre-clinical models of various aging and pathologic conditions (osteoporosis, osteoarthritis, and arthritis) as well as models for tissue regeneration and bone fracture repair as well as high-throughput skeletal phenotyping. Pre-clinical *in vivo* imaging methods are now routinely used as a screening tool in drug discovery and development processes. The availability of sophisticated software for analyses and visualization of these images has made micro-CT a very versatile research tool for the pre-clinical laboratory and promises to do the same for clinical research.

9:15

5aBB4. Bone diagnostics using dual frequency ultrasound measurements. Jukka S. Jurvelin, Janne Karjalainen, Ossi Riekkinen, Markus Malo, Hanna Isaksson, and Juha Tyräs (Dept. of Phys. and Mathematics, Univ. of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland, jukka.jurvelin@uef.fi)

Diagnosis of osteoporosis is made at skeletal sites composed mainly of trabecular bone. Calcaneus has been the first location for through transmission (TT) ultrasonic measurements of trabecular bone. Similarly as with the DXA, the best prediction of the hip fractures could be obtained by making the measurements at hip. To realize axial pulse-echo (PE) ultrasound measurements, e.g., in hip we have introduced the dual frequency ultrasound (DFUS) technique to minimize the effects of soft tissues overlying the bone and to correct the PE-parameters [Riekkinen *et al.*, Ultrasound Med. Biol. **34**, 1703–1708 (2008)]. Based on our experimental measurements at frequencies of 2.25 and 5.0 MHz and finite difference simulations, the DFUS technique detects minor changes in bone density, despite variable composition of soft tissue. For TT-geometry, the composition of the interfering soft tissues may also be solved with the DFUS, e.g., by measuring the reflection from the bone surfaces at both sides of the bone. The measurements with phantom materials indicated that accuracy of TT-parameters, such as broadband ultrasound attenuation, were significantly improved by DFUS. Potentially, the DFUS could also benefit the *in vivo* TT-measurements of trabecular bone. [Support from Academy of Finland is acknowledged.]

Contributed Papers

9:30

5aBB5. Comparison of conventional ultrasonic phase velocity and attenuation measurements of cancellous bone to estimates obtained using Bayesian probability theory. Christian C. Anderson (Washington Univ. in St. Louis, 1 Brookings Dr., St. Louis, MO 63130), Michal Pakula (Kazimierz Wielki Univ., Bydgoszcz, Poland), Pascal Laugier (Université Pierre et Marie Curie, Paris, France), G. Larry Bretthorst, Mark R. Holland, and James G. Miller (Washington Univ. in St. Louis, St. Louis, MO)

Cancellous bone supports the propagation of two compressional waves, commonly referred to as fast and slow waves, which can overlap in the acquired time-domain data. When such data are processed using conventional techniques, interference effects may cause artifacts to be present in the speed of sound and broadband ultrasound attenuation (BUA) measurements. Bayesian probability theory is a proposed method for extracting the ultrasonic properties of the individual fast and slow waves, even in the presence of substantial interference. In the current study, data were acquired *in vitro* at sites on a cancellous bone, and the acquired signals were used as input to a Bayesian probability approach for estimating parameters characterizing the fast and slow waves. For comparison, the signals were also processed using conventional methods. Although the conventionally obtained phase velocity and BUA frequently exhibited anomalous features, including negative dispersion, those estimated using probability theory exhibited no aberrant behavior. Hence, material properties obtained using Bayesian probability theory may provide more reliable estimates of bone quality and fracture risk than the apparent properties derived from the non-causal mixed mode signals. [Work supported by NIH R01HL40302 and R01AR057433, NSF CBET-0717830.]

9:45

5aBB6. Transfer functions for one-dimensional bone models. James L. Buchanan (Dept. Mathematics, US Naval Acad., 572C Holloway Rd., Annapolis, MD 21401, jlb@usna.edu), Robert P. Gilbert (Univ. of Delaware, Newark, DE 19711), and Miao-Jung Ou (Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

A system for calculating wave pulses through one-dimensional layered media via a series of transfer functions is developed. This system obviates the need for numerical approximation beyond the use of discrete Fourier transforms, whence it is not necessary to deal with the poor numerical conditioning of the linear systems that arise in multi-layered models. A one-dimensional model of bone as a cancellous segment between two cortical segments insonified by an ultrasonic pulse is considered. For the cancellous

segment the Johnson–Dashen–Koplik modification of Biot's model for a poroelastic medium is used. The cortical segments are modeled both as dense fluids or porous media and the results compared.

10:00—10:15 Break

10:15

5aBB7. Fabric dependence of wave attenuation in anisotropic poroelastic media. Luis Cardoso and Stephen Cowin (Dept. of Biomedical Eng., CCNY, New York, NY 10031, scowin@earthlink.net)

The governing equations for wave motion in the linear theory of anisotropic poroelastic materials have been developed and extended to include the dependence of the constitutive relations upon fabric. Fabric is a quantitative stereological measure of the degree of structural anisotropy in the pore architecture of a porous medium. This anisotropic poroelastic theory is then applied to cancellous bone architecture to study the role of both porosity and fabric on the wave attenuation due to reflection, absorption, and scattering. The analysis of the fast and slow wave attenuation is performed as a function of the porosity of the medium and the frequency of the wave. This study demonstrates that the attenuation ratio between fast and slow waves strongly depends on the porosity and fabric of the trabecular structure. Wave attenuation predictions from this model confirm our previous experimental observations indicating that either the fast or slow wave may be the predominant wave mode in trabecular bone depending on the bone architecture. [This work was supported by grants from NSF (NSF/CUNY AGEP 0450360, NSF-MRI 0723027, and PHY-0848491). The authors also acknowledge the support from PSC-CUNY Research Award Program (PSC-CUNY-60014-38-39, 69486-00 38, and PSCREG-39-1103) and NIH 1SC2AG34198-1A1.]

10:30

5aBB8. Fabric dependence of tortuosity in poroelastic wave propagation. Stephen Cowin and Luis Cardoso (Dept. of Biomedical Eng., CCNY, New York, NY 10031, scowin@earthlink.net)

The governing equations for wave motion in the linear theory of anisotropic poroelastic materials have been developed and extended to include the dependence of the constitutive relations on fabric. Fabric is a quantitative stereological measure of the degree of structural anisotropy in the pore architecture of a porous medium. With the addition of the second order symmetric tensor fabric variable, the formulation of wave motions in the poroelastic theory is consistent with the presentations of Biot and later authors. In this work we show that tortuosity is a function of the fabric tensor. This relationship between tortuosity and fabric is employed to provide a geometric understanding of tortuosity in terms of pore space architecture. This geometric understanding is then applied to cancellous bone architecture. [This work was supported by grants from NSF (NSF/CUNY AGEP 0450360, NSF-MRI 0723027, and PHY-0848491). The authors also acknowledge the support from PSC-CUNY Research Award Program (PSC-CUNY-60014-38-39, 69486-00 38, and PSCREG-39-1103), and NIH 1SC2AG34198-1A1.]

10:45

5aBB9. Inverse problem of recovery of poroelastic parameters of cancellous bone by inversion of transmitted ultrasonic data. Zine El Abiddine Fellah, Erick Ogam (Laboratoire de Mecanique et d'Acoustique, UPR 7051 CNRS, 31 Chemin Joseph Aiguier, 13402 Marseille, France), Catherine Masson (INRETS-UMR T 24 LBA UFR de Medecine, 13916 Marseille Cedex 20, France), and Robert Gilbert (Univ. of Delaware, Newark, DE 19716)

This investigation addresses the inverse problem (IP) of the retrieval of the macroscopic poroelastic parameters of cancellous bone using transmitted ultrasonic waves. The IP is solved by minimizing the objective functional (OF), which is a measure of the discrepancy, in the least squares sense, between the trial and measured data pertaining to the transmitted USW. In this study the Biot–Johnson–Koplik was employed as the interaction model (IM). A close examination of the OF computed using this IM and measured ultrasonic data transmitted through the cancellous bone showed that it exhibited several local minima and maxima. Most off-the-shelf iterative nonlinear least squares optimization algorithms (such as the Levenberg–Marquadt or Nelder–Mead simplex methods) could not retrieve a nonambiguous solution. A simple method for resolving this problem without ambiguity by using two distinct transducer central frequencies is proposed. Finally the recovered acoustic parameters are compared with those obtained using classical methods.

11:00

5aBB10. Frequency dependence of phase shifts in human calcaneus *in vitro*. Keith A. Wear (FDA Ctr. for Devices and Radiological Health, Bldg. 62, Rm. 3108, 10903 New Hampshire Blvd, Silver Spring, MD 20993)

If dispersion in a medium is weak and approximately linear with frequency (over an experimental band of frequencies), then it can be shown that the constant term in a polynomial representation of phase shift as a function of frequency can produce errors in measurements of phase velocity differences in through-transmission, substitution experiments. A method for suppressing the effects of the constant phase shift in the context of the single-wave-model was tested on measurements from 30 cancellous human calcaneus samples *in vitro*. Without adjustment for constant phase shifts, the estimated phase velocity at 500 kHz was 1516 +− 6 m/s (mean +− standard error), and the estimated dispersion was −24 +− 4 m/sMHz (mean +− standard error). With adjustment for constant phase shifts, the estimated mean velocity decreased by 4–9 m/s, and the estimated magnitude of mean dispersion decreased by 50%–100%. The average correlation coefficient between the measured attenuation coefficient and frequency was 0.997 +− 0.0026 (mean +− standard deviation), suggesting that the signal for each sample was dominated by one wave.

11:15

5aBB11. Computing bone fragility using transient ultrasonic waves. Robert P. Gilbert and Philippe Guyenne (Dept. Mathematical Sci., Universirty of Delaware, 317 Ewing Hall, Newark, DE 19711)

In this paper, we determine bone fragility of cancellous bone *in vivo* by using a simple one-dimensional model of a muscle-cortical bone-cancellous bone sample, by measuring the refracted sound when the sample is exposed to impulses from a transducer. From these data, the bone parameters can be determined. Such a procedure is referred to as solving an inverse problem. A transducer is placed on one side of the skin of the member to be interrogated and a receiver is place diametrically opposed in contact with the member. The muscle will be modeled as an elastic material, whereas the the corrical and cancellous bone may be modeled by modified Biot equations albeit with different porosities. However, because of the known frequency dependence of the viscous term, we propose that the standard Biot equations, which are not well posed, be replaced by adding a convolution term. The form of the kernel $b(t)$ in the convolution integral is suggested by frequency domain representation of the dissipation term in the Biot–Johnson–Koplik–Dashen model.

11:30

5aBB12. Ultrasonic characterization of cancellous bone using three models of trabecular structure. Jerzy Litniewski, Andrzej Nowicki, and Janusz Wojcik (Dept. of Ultrasound, Inst. of Fundamental Tech. Res., Pawinskiego 5b, 02-106 Warsaw, Poland, jlitn@ippt.gov.pl)

The semi-empirical scattering models of trabecular bone were developed and examined for their abilities to mimic the frequency dependent backscattering coefficient measured in the cancellous bone. In the simulation of the bone, rf echoes the real properties of the bone and experimental conditions were taken into account. Tree types of trabeculae mimicking scatterers were considered. First, the bone consisted of cylinders with varying thickness (Gamma distributed) within the population was assumed. The next two cases accounted for the contribution of thick and thin trabeculae to the total backscattered signal. The second model assumed existence of two populations of the cylindrical scatterers significantly differing in the average value of Gamma distributed diameters. Finally, the mixed model composed of thick and thin trabeculae modeled, respectively, by cylindrical and spherical scatterers was examined. The last selection resulted from the similarity found between scattering on small sphere and finite cylinder. Calculated echoes demonstrated the usefulness of the mixed model. Frequency dependence of backscattering coefficient agreed well with the experimentally de-

termined dependences. The study also showed that the amplitude histograms calculated using demodulated rf echoes deviate from the Rayleigh distribution when the variation in scatterers' diameters increase.

11:45

5aBB13. Quantitative evaluation of articular cartilage using a minimally invasive ultrasound method. Tuomas Viren, Simo Saarakkala, Jukka S. Jurvelin, and Juha Tyräs (Dept. of Phys. and Mathematics, Univ. of Eastern Finland, P.O. Box 1627, 70210 Kuopio, Finland, tuomas.viren@uef.fi)

Quantitative ultrasound imaging is a potential method for the evaluation of the integrity of articular cartilage. A quantitative minimally invasive intra-articular ultrasound (IAUS) technique has been developed for *in situ* and *in vivo* diagnostics of articular cartilage [Virén *et al.* Ultrasound Med. Biol. **35**, 1546–1554, (2009)]. In IAUS, a high-frequency (40 MHz) intravascular ultrasound device is used and reflection coefficient (R), integrated reflection coefficient (IRC), apparent integrated backscatter coefficient (AIB), and ultrasound roughness index (URI) are calculated for cartilage. A set of experiments was conducted for the validation of the technique. For example, significant differences were observed in ultrasound parameters between intact, spontaneously and surgically repaired rabbit articular cartilage. The parameters revealed abnormal surface roughness and internal collagen structure by decreased R and IRC and increased values of URI and by increased AIB. Furthermore, the poor integration of the repair and surrounding native tissue was visible in the ultrasound images. As a reference, histological images revealed that repaired tissue exhibited abnormally organized collagen network, rough surface and lower collagen, and proteoglycan contents. Based on the present experimentation using IAUS, tissue structure and properties, not diagnosable by present clinical techniques, may be quantitatively evaluated with the minimally invasive ultrasound technique.

FRIDAY MORNING, 23 APRIL 2010 LAUREL A/B, 8:00 TO 11:15 A.M.

Session 5aEA

Engineering Acoustics: Electret Condenser Microphones

Allan J. Zuckerwar, Cochair
4909 Camberley Cir., Williamsburg, VA 23188

Qamar A. Shams, Cochair
NASA Langley Research Ctr., 4 Langley Blvd., Mail Stop 238, Hampton, VA 23681

Chair's Introduction—8:00

Invited Papers

8:05

5aEA1. Electret condenser microphones. James E. West (Dept. of Elec. and Comput. Eng. and Dept. of Mech. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21231, jimwest@jhu.edu)

For more than 4 decades electret microphones have dominated most transducer applications in the audio frequency range for three main reasons, linearity, simplicity, and they are inexpensive. E. C. Wente invented the condenser microphone in 1916 searching for a replacement for the carbon microphone used in most telephones, but it was impractical until the high-voltage battery was replaced with a permanently charged polymer. Electret microphones have been the workhorse for most microphone applications including microphones for professional sound level measurements, telephones, hearing aids, professional audio recordings, games, and toys. Recent worldwide estimates of the electret microphone market suggest that over 2 billion units are made annually. Electret condenser microphones are available in a variety of sizes from very small (a few millimeters in diameter) to about 2 in. in diameter. The use of multiple microphone elements in arrays and the use of digital signal processing provide interesting prospects for improving signal-to-noise. A variety of microphone arrays and some important historical facts will be discussed in this talk.

8:25

5aEA2. Piezoelectret microphones. Gerhard M. Sessler and Joachim Hillenbrand (Dept. of Elec. Eng., Univ. of Technol., Merckstrasse 25, 64283 Darmstadt, Germany, g.sessler@nt.tu-darmstadt.de)

Piezoelectret microphones, a recently described new class of electret microphones, are based on the strong piezoelectric effect of internally charged, cellular polymer films. Such piezoelectret films have a large number of internal voids within the polymer. The upper and lower walls of these voids are oppositely charged. Films of this kind, when metallized on both sides, show high-piezoelectric activity and can be directly used as microphones. Recent work on these transducers has concentrated on stacking of piezoelectret films to improve sensitivity, on directional microphones, and on investigations of new film materials with improved charge stability at elevated temperatures. Stacked transducers have sensitivities proportional to the number of piezoelectret layers. With six layers, values of approximately 15 mV/Pa can be achieved. Directivities include omnidirectional, bidirectional, and cardioid patterns as well as dimensional characteristics. The new materials are based on fluoropolymers. These are, as opposed to the originally used polypropylene, not available in suitable cellular form. However, useful films can be made by fusing thin fluorocarbon films such that an array of small air voids is created between the layers. Transducers consisting of charged films of this kind withstand temperatures of 90 °C without significant loss of sensitivity.

8:45

5aEA3. Temperature characteristics of electret silicon microphones. Yoshinobu Yasuno (Appl. Piezoelectricity Div., Kobayasi Institut of Physical Res., 3-20-41, Higashi-motomachi, Kokubunji, Tokyo 185-0022, Japan, yasuno@kobayasi-riken.or.jp)

The electret condenser microphone (ECM) has become an important component for various consumer equipment systems because of its stable sensitivity and frequency characteristics and its success in achieving small size and high sensitivity. A previous report described the method to design a microphone with stable temperature characteristics. The quality of the silicon material for the diaphragm was pointed out as important for improving robustness. Recently, microphones using micro-electro-mechanical systems (MEMS) have been practically applied and widely used for mobile equipment, such as cellular phones. The major reason for adopting MEMS is that a re-flow soldering process can be used in production, in addition to having desirable features such as being small and thin. However, almost no electret types have been commercialized since a guaranteed 300°C heat resistance is required. This report presents a new trial electret silicone microphone that contains a SiO_2 thin-film electret. The report also evaluates the heat resistance of a SiO_2 electret, the temperature stability of a silicon diaphragm, and the temperature change in sensitivity and frequency characteristics compared with those of a conventional ECM and the microphone for measurement.

9:05

5aEA4. Prepolarized measurement microphone properties: 30 years of experience. Johan Gramtorp and Erling Sandermann Olsen (Brüel & Kjær Sound & Vib. Measurement A/S, Transducers R&D, Skodsborgvej 307, 2850 Nærum, Denmark jgramtorp@bksv.com)

30 years ago, Brüel & Kjær launched the world's first prepolarized condenser measurement microphones. These microphones combined the high-performance construction of the normal measurement microphones with a polarized electret layer placed on the back electrode. Since then, prepolarized microphones have gained an increasing share of the measurement microphone market. Prepolarized microphones are flexible in use because they do not need the external polarization, in particular, they can be used with constant current supply preamplifiers (Deltatron, ICP) that are widely used for many purposes in the industry. In order to be used for measurement purposes, the microphones and thereby the electret polarization must be highly stable and reliable. That is indeed achievable, and that is the main emphasis of this paper. A short historical overview of the development of the Brüel & Kjær prepolarized microphone types is given. The properties of prepolarized microphones are discussed and compared to those of externally polarized microphones. These will include stability with respect to time, temperature, humidity and the chemical environment, as well as distortion and dynamic range.

9:25

5aEA5. Electret condenser microphones in hearing aids. Daniel Warren, David Schafer, and Kiran Konde (Knowles Electron., 1151 Maplewood Dr., Itasca, IL 60143, daniel.warren@knowles.com)

As sensors small, high-gain devices, hearing aid microphones must be tiny, have high signal-to-noise ratios, be relatively insensitive to vibration from the nearby speaker, and operate at low hearing aid battery voltages. The subminiature backplate-electret condenser microphone was developed in the early 1970s specifically for hearing aids, with a very lightweight mechanical system for lower vibration sensitivity and somewhat improved signal-to-noise compared to previous designs. Although the design has been refined over several generations, the basic technology of electret films has proven to be very robust over the last 40 years, enabling continual size reductions while maintaining high signal-to-noise ratios. Noise generation and vibration sensitivity continue to be competitive dimensions for hearing aid microphones. Some of the mechanisms for noise generation and vibration pickup are intrinsic to the design of these microphones. Other mechanisms may be caused by interaction between the microphones and the hearing aid system of which they are a part. In this presentation, we will discuss both types of mechanisms and some design guidelines for minimal noise application of microphones in hearing aids.

9:45—10:00 Break

10:00

5aEA6. Ringing the changes: The potential for push-pull electret transducers in cell phones. Tim Mellow (Devices R & D, Nokia, Summit Ave., Farnborough GU14 0NG, UK, tim.mellow@nokia.com) and Leo Kärkkäinen (Nokia Res. Ctr., 00180 Helsinki, Finland)

Electret microphones are very common in mobile devices, but are invariably single-ended in configuration. For moderate dynamic conditions this is not a problem, but under high-sound pressure levels, better linearity can be obtained using a push-pull arrangement. Recent progress in nano-porous electret materials also opens up possibilities for electret loudspeakers. Such loudspeakers would offer potential benefits such as high-conversion efficiency, no magnetic field, and an ultra-thin package. However, there are product integration challenges. The authors describe a simulation model of a push-pull electret transducer and discuss the conditions under which it is linear. It turns out that the relationship with a non-electret electrostatic transducer is relatively simple.

10:20

5aEA7. Airframe noise flight test measurements using an electret condenser microphone array. William M. Humphreys, Jr., David P. Lockard, and Mehdi R. Khorrami (NASA Langley Res. Ctr., Hampton, VA 23681)

An extensive collection of airframe noise flight test data has been acquired for a regional jet class of aircraft using a ground-based electret condenser microphone phased array system. The measurements were conducted at the Wallops Flight Facility for the NASA Aeronautics Research Mission Directorate Fundamental Aeronautics Program. Gulfstream G450 and G550 aircraft were used as the test beds for the study. The phased array employed for the measurements consisted of a spiral clustering of 167 Panasonic WM-61A quarter-inch, omni-directional electret microphones. The innermost 49 sensors of the array were flush mounted on an aluminum plate with the remaining sensors flush mounted on individual 16-in. circular Plexiglas platters. The aluminum plate and platters were deployed on the overrun area of runway 4 at Wallops. Airframe noise measurements were performed with the aircraft flying a racetrack pattern simu-

lating the approach path to landing as it passed over the array. A distributed data acquisition system was used to acquire simultaneous time histories from the array and was synchronized with instrumentation onboard the aircraft. This presentation will discuss the design of the phased array, the calibration and deployment of the microphones, a sampling of the measurements obtained from the system, and lessons learned from the deployment.

10:40

5aEA8. A method for free-field transfer calibration of electret microphones at ultrasound frequencies in air using a stable broad band ultrasound source. Angelo J. Campanella (Campanella Assoc. and Acculab, 3201 Ridgewood Dr., Hilliard, OH 43026, a.campanella@att.net)

A rotating turbulence screen wheel sound source, now in use in the air ultrasound instrument industry for product quality control, provides a steady source of broad band ultrasound. This source is broad band through at least 100 kHz. It is calibrated as a secondary standard, using a one-quarter inch condenser microphone, grid removed, normal incidence, and corrected for free-field conditions per the microphone manufacturer free-field response. Sound pressure amplitude is measured at a distance of 500 mm and from 1 to 100 kHz in 1-kHz FFT bins. Calibration site sound absorption by humid air is calculated to correct the emitted sound pressure values to be reinserted for the user test site air temperature and humidity. Transfer calibrations of two electret condenser microphones are shown. Microphone phase response measurement methods will also be investigated and discussed.

Contributed Paper

11:00

5aEA9. Acoustic field of air-coupled transducers made with electroactive polymer foams. Louis Satyanarayan and Yves Berthelot (Woodruff School of Mech. Eng., Georgia Tech Lorraine, GT-CNRS UMI 2958, 2 rue Marconi, 57070 Metz, France)

Electroactive films made of soft cellular polypropylene (PP) foams containing electric charges have high piezoelectric coupling coefficient and good matching impedance with air over a broad frequency range. They offer interesting capabilities in the design of new air-coupled ultrasonic transducers. A prototype transducer has been designed and tested in our laboratory. The amplitude displacement on the surface of the transducer was measured by laser vibrometry by vibrating the PP foams with a continuous wave sinusoidal excitation. The transducer was mounted with a needle-type PVDF hydrophone on a high-precision polar C-scan apparatus. The receiving hydrophone was translated along and across the axis of the source transducer and data were recorded. A quantitative characterization of the acoustic field of the transducer was carried out by plotting the on-axis and cross-axis variations in the pressure field, at various separation distances between the source and the receiver in pulsed transmit-receive mode. The results indicate that simple and relatively efficient air-coupled ultrasonic transducers can be made with electroactive polymer foams. [Work supported by CNRS and by the Conseil Rgional de Lorraine.]

FRIDAY MORNING, 23 APRIL 2010 KOZMOS, 8:20 A.M. TO 12:00 NOON

Session 5aMU

Musical Acoustics: Extended Instrument Techniques

Jonas Braasch, Chair

Rensselaer Polytechnic Inst., School of Architectue, 110 8th St., Troy, NY 12180

Invited Papers

8:20

5aMU1. Instrumental analysis of extended saxophone techniques for live electronics. Jonas Braasch and Doug Van Nort (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, braasj@rpi.edu)

The development of automated music transcription systems focuses predominantly on polyphonic musical instruments. At the same time, the analysis of a monophonic instrument is usually much simpler wherein pitch, loudness, and duration of individual notes may be tracked robustly. When using extended techniques, however, many more parameters than the aforementioned three can be meaningful for the performed music. This paper explores the challenges that extended techniques pose for music recognition systems using the example of the saxophone. The goal is to correctly identify extended techniques over the whole range of the instrument, including subtones, multiphonics, growl, and other voice-enhanced tones, as well as tones where the reed is supported by the lower teeth. The feature analysis is based on cepstrum, spectral moments, pitch, and roughness, among other features. A hidden-Markov model is used to recognize the trajectory of the various extended techniques based on the given feature space. Finally, it is demonstrated how the recognizer can be integrated into an intelligent live electronics system to control its parameters. For example, the characteristics of a virtual acoustic enclosure (room size, reverberation time, etc.) can be adapted this way. [Work supported by NSF 0757454.]

8:40

5aMU2. Pitch bending and higher-mode reed vibration in mechanically blown free reed instruments. James P. Cottingham (Dept. Phys., Coe College, 1220 First Ave. NE, Cedar Rapids, IA 52402, jcotting@coe.edu)

Pitch bending in the harmonica, in which manipulation of vocal tract resonances plays an essential role, has long been a common practice. Pitch bending in free reed instruments with mechanically driven air supplies, such as the reed organ and the accordion, is a different matter. At least two methods of pitch bending in such instruments have been studied and demonstrated. The first, likely to be musically useful only in certain special circumstances, involves partial opening of the pallet valve combined with variations in blowing pressure. The second, recently described and implemented in the accordion by Tonon [J. Acoust. Soc. Am. **126**, 2217 (2009)], involves modifying the construction to include a resonating chamber in addition to the standard reed chamber. It is also possible to construct air-blown free reed instruments in which the reed vibrates mainly in the second transverse mode. In one case the reed is coupled with a pipe resonator that provides a suitable mismatch in frequencies with the fundamental frequency of the reed. In another case the second or higher-mode vibration of the air-driven reed can be mechanically induced. The musical possibilities of these hypothetical reed-pipe instruments have yet to be explored.

9:00

5aMU3. Extended techniques of the contrabass. Michael T. Bullock (15 Channel Ctr. St., Apt. 403, Boston, MA 02210, michaeltbullock@gmail.com)

Contrabass improvisers, composers, and performers of contemporary contrabass repertoire rely on a broad vocabulary of extended contrabass techniques. This vocabulary is continuously developed primarily through a kinesthetic engagement with the acoustics of the contrabass, and many new discoveries are made in the act of playing. This presentation will include live performance, by the presenter, of a series of extended techniques on the contrabass. Spectrograms and video clips will be used to illustrate how these techniques are related to the unique acoustic properties of that instrument. Principles demonstrated will include the following: (1) *Overtone bowing and fingering*: The low fundamental register of the contrabass means that each string has a many upper overtones within the audible range. Using various bowing and fingering techniques that exploit these upper partials, the timbre and pitch of a string can be radically altered and continuously transformed while playing. (2) *Activating the contrabass body*: Striking the contrabass body with hands and soft objects, rubbing the instrument's back with the fingers, and manipulating feedback through amplification are ways to activate the complex resonances of the contrabass. [Thanks to Professor Jonas Braasch of the RPI School of Architecture.]

9:20

5aMU4. Extending and abstracting sitar acoustics in performance. Curtis R. Bahn (Dept. of the Arts, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180-3590, crb@rpi.edu)

There is a very close relationship between the acoustic design of instruments of the Hindustani classical music tradition and the structure of raga. The sitar and other related string instruments have sympathetic strings which are tuned to prominent notes of a raga. These strings resonate sympathetically in performance, creating a rich and hauntingly beautiful sonority. In creating a computer extended electronic sitar, the first step was to model and extend this resonant quality of the instrument, allowing for new tunings and extended resonances not possible with a traditional instrument. The extended sitar also uses performance analysis and a variety of sensors to control various signal processing techniques that further extend the quality of the instrument. This sonic analysis and subsequent sound synthesis/computer generated compositional schemes relate directly to structures of raga. This presentation will detail the development and performance of the computer extended sitar in relationship to traditional raga structure and in new contexts not directly related to traditional notions of the sitar.

9:40

5aMU5. Simulating, mirroring, morphing, and blurring acoustic space as an extended instrumental technique. Pauline Oliveros (Dept. Arts, Rensselaer Polytechnic Inst., 110 8th St., Troy NY 12080)

The paradigm for musical performance is the sound of musical instruments in a fixed acoustical space. Ordinarily space does not shift and change during performances. An ideal space for performing classical composers such as Mozart is the Shaffy Theater in Amsterdam with a relatively short reverberation time 1.5 s. Instruments couple to the space, and space is part of the instrument and vice versa. Players develop special relationships with favorite spaces. After experiencing many different types of acoustics this leads to the desire to change the space during performance in order to hear the different qualities that emerge from spatial coupling. Spatial progressions can be made so that rather than moving sound in space, space can be moved around the sound and produce a variety of qualities.

10:00—10:15 Break

10:15

5aMU6. Extending the acoustic ensemble through spectral and temporal transformations in real-time. Doug Van Nort (Dept. Architectural Acoust. and Dept. Arts, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, vannod2@rpi.edu)

The paradigm of live performance mixing acoustics and electronics has predominantly focused on simple background "tape music," human players performing highly structured sample-based music (e.g., using the ABLETON LIVE software), or reactive systems that respond to player qualities such as timing, pitch, and so on. In this talk I will present my approach to improvised "laptop performance" that focuses on the transformation of acoustic players in real-time. Rather than simply altering the acoustic content in the manner of an effect processor, the goal is to capture notes and phrases in short-term memory and to re-articulate the material so that it presents a new gestural inflection and timbral content that can be completely novel or suggestive of other players' sound. The system presented utilizes

a hybrid system combining spectral analysis and feature extraction with block-based temporal processing and a feedback delay network. The interaction paradigm of "scrubbing" the intermediate time/frequency representation is used to generate the ultimate output. The result in an ensemble context is an extended palette that can "keep up" with the musical dialog while eliciting the subtle textural qualities of acoustic players. [This work was supported by NSF Grant 0757454 and CIRMMT/McGill University Fellowships.]

10:35

5aMU7. The digital audio player as aleatory musical instrument. Adam Di Angelo (SIA Acoust., 75 Rockefeller Plaza, 26th Fl., New York, NY 10019, diangelo@gmail.com)

The digital audio file player is an ideal instrument for the performance of aleatory music. The near-instantaneous seek time coupled with random track playback yields the possibility of nonlinear musical arrangements. In this is the potential to create unique and unrepeatable listening experiences—music that constantly renews and reinvents itself, free from the static limitations of the recorded medium. This study examines the history of chance as a compositional and performance tool from 18th century musical dice games to contemporary stochastic methods. It presents examples of works which are designed to be structurally random and outlines the process of employing the digital audio player in the production of variable music. Through such considerations, the role of the conventional mp3 player can expand from passive playback device to become an engine integral to a new kind of listening.

Contributed Paper

10:55

5aMU8. The physics of pitch bending on the vibraphone. Randy Worland (Dept. of Phys., Univ. of Puget Sound, 1500 N. Warner, Tacoma, WA 98416, worland@pugetsound.edu)

The vibraphone is a member of the keyboard (or mallet) percussion family, with tuned aluminum bars typically spanning three chromatic octaves (F3–F6). Extended performance techniques include the use of harmonics and bowing of bars, as well as the pitch bending effect discussed here. After first striking the middle of a bar with a soft mallet in the normal manner, the pitch bend is obtained by pressing a hard mallet onto the bar at a nodal point and then sliding it away from the node as the note sustains. The audible result is a descending pitch, typically of about one semitone. Experimental data on frequency versus location of the hard mallet along the bar are presented. The mass, hardness, and orientation of the sliding mallet relative to the bar are discussed. Interferograms of the lower frequency modes of the bar are also shown. Results are interpreted using a mass perturbation model.

11:10—11:15 Break

11:15—12:00
Participants will give informal demonstrations of several of the techniques discussed.

FRIDAY MORNING, 23 APRIL 2010 GRAND BALLROOM III/IV, 8:00 TO 9:45 A.M.

Session 5aPAa

Physical Acoustics: Theoretical Models

Yves H. Berthelot, Chair
Georgia Inst. of Technology, School of Mechanical Engineering, Atlanta, GA 30332-0405

Contributed Papers

8:00

5aPAa1. Green's function for the scattered field near the surface of a dissipative elastic medium. Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

The problem of scattering in bounded elastic media is relevant to medical ultrasound imaging and detection of buried objects. A method was developed that takes into account the propagation of longitudinal, shear, and Rayleigh waves in an elastic medium. The approach is based on solving the inhomogeneous momentum equation for the Green's function. The forcing function is modified to accommodate different kinds of excitation that are important for shear wave imaging and seismic wave generation. The solution for the Green's function is presented as a Fourier integral, and coefficients in the Fourier expansion are calculated by applying boundary conditions on the solid surface and matching conditions in the plane parallel to the surface that contains the source. The method permits the scattered bulk and surface waves to be distinguished from one another. The Green's function has been presented previously in both the frequency and time domains for a lossless medium. The model is generalized here for elastic media with losses. Besides reducing the amplitude of the response, losses also impair ability to distinguish between the bulk and surface modes of propagation. [Work supported by NIH DK070618.]

8:15

5aPAa2. Image method for acoustic reflections from a wedge. Ambika Bhatta, Max Denis, Mira Raspopvic, and Charles Thompson (ECE Dept., Univ. Mass. Lowell, 1 Univ. Ave. (CACT), FA 203, Lowell, MA 01854)

In this paper we will present the analysis of reflected waves of a point source from a two dimensional wedge. Exact method of images is used to calculate the position and strength of image sources. It is shown that the traditional fractional order Bessel function series for the total pressure can be replaced by quadratures of elementary functions.

8:30

5aPAa3. Transition matrix (T-matrix) of circular cylinders embedded in a viscous fluid. Edgar Reyes-Ayona, Daniel Torrent, and Jose Sanchez-Dehesa (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de vera s.n., ES-46022 Valencia, Spain)

The present study concerns the scattering of a plane sound wave by a cylindrical scatterer of circular cross section. An infinitely long cylinder embedded in a viscous fluid is here considered. The linearized equations of viscous fluid and the scalar and vector potentials are employed to obtain the T-matrix for two different types of material cylinders: (i) a rigid cylinder and (ii) a fluidlike cylinder with losses. Both cases are also studied and discussed in the limit of long wavelengths. [Work supported by MICIIN of Spain and ONR.]

8:45

5aPAa4. A numerical model for acoustic scattering from an inhomogeneous medium. Max Denis, Jing Tsui, Jessica Piper, Kavitha Chandra, and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 Univ. Ave., Lowell, MA 01854, charles_thompson @uml.edu)

In this work, a numerical method for evaluating the scattering of acoustic waves from an inhomogeneous medium is presented. The internal resonant and multiple scattering effects are modeled by the Padé approximants of a perturbation expansion of the contrast properties between the medium constituents. Due to the local divergence in the Padé approximant solution, a local expansion method about any convergent solution is performed. This allows one to extend the numerical model to higher frequencies. The results show good agreement for geometrically simple scatterers. Based on the simulations, analyses are made of inhomogeneous media micro-structural effects on the scattered field. Of particular interest is the frequency dependence relationship.

9:00

5aPAa5. Single-channel time reversal with Lanczos iterations. Zachary J. Waters (Naval Res. Lab., Code 7130, Washington, DC 20375-5320) and Paul E. Barbone (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Recent investigations of array-based time reversal techniques employing Lanczos iterations have demonstrated convergence properties superior to equivalent methods using power iterations [A. A. Oberai *et al.*, "Efficient time reversal by Lanczos iterations," J. Acoust. Soc. Am. **123**, 3596 (2009)]. In numerical scattering simulations, time reversal techniques developed for a single-channel transducer are applied to study echoes from a thin-walled spherical shell. When power iterations are employed, a narrowband waveform corresponding to the dominant scattering mode of the target is selected. Lanczos iterations converge to this waveform in fewer transmissions and simultaneously identify additional higher-order target scattering features. The performance of both techniques is studied in the presence of varying levels of stochastic and deterministic noise. [Work supported by the Office of Naval Research.]

9:15

5aPAa6. Modeling of acoustic propagation to capture finer-scale features of bounding objects. Sandra L. Collier, W.C. Kirkpatrick Alberts, II, and John M. Noble (U.S. Army Res. Lab., RDRL-CIE-S, 2800 Powder Mill Rd., Adelphi, MD 20783, scollier@arl.army.mil)

One of the limiting factors of performing high-resolution modeling of outdoor sound propagation in the time domain is the treatment of the boundary conditions. Impedance ground boundary conditions have been successfully modeled in the time domain for a variety of porous ground types via the use of recursive convolution methods; however, the applications were for a flat ground. The modeling of sound propagation around buildings (or other obstacles) is usually accomplished by drastically simplifying the building geometry, thereby failing to account for fine-scale, or even some larger-scale, details. Experiments conducted at a fire training facility found that external building features, such as facets and fire escapes, can effect the sound field significantly enough so that standard acoustic diffraction models fail to accurately predict the sound field. Here we develop numerical models that allow us to capture the effects of finer-scale features. Such treatments are necessary for the development of source localization algorithms in urban or other multifarious environments.

9:30

5aPAa7. Energy flux streamlines vs the alternatives for the visualization of energy coupling inside and outside the surface of an ensonified spherical acoustic lens: Preliminary results. Cleon E. Dean (Dept. Phys., P.O.B. 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean@georgiasouthern.edu) and James P. Braselton (Georgia Southern Univ., Statesboro, GA 30460-8093)

Energy flux streamlines are compared with alternative presentations of the energy flux vector field to visualize energy coupling inside and outside the surface of an insonified spherical acoustic lens. The emphasis is on the energy flux streamline as a natural bridge between necessarily approximate ray solutions and full-fledged wave solutions. The present work is the wave solution for the experimental work performed by Kendez C. Parker and Cleon E. Dean also being presented at this meeting that primarily used ray methods in its theoretical analysis; thus a comparison between energy flux representations and ray analysis results is made. Difficulties in the interpretation of the energy flux streamline representation due to the spherical geometry of the scatterer are noted.

Session 5aPAb

Physical Acoustics: Special Topics and Applications in Physical Acoustics

Andi G. Petculescu, Chair
Univ. of Louisiana, Lafayette, Dept. of Physics, P.O. Box 44210, Lafayette, LA 70504

Contributed Papers

10:15

5aPAb1. Transmission of acoustic-gravity waves through air-water interface. Iosif M. Fuks (ZelTechnologies LLC and NOAA/Earth System Res. Lab., Mail Code R/PSD99, Boulder, CO 80305-3328) and Oleg A. Godin (Univ. of Colorado, Boulder, CO 80305-3328, oleg.godin@noaa.gov)

It was demonstrated recently that air-water interface, which is usually an almost perfect reflector of acoustic waves, becomes anomalously transparent and the power flux in the wave transmitted into air increases dramatically, when a compact sound source in water approaches the interface within a fraction of wavelength [O. A. Godin, Phys. Rev. Lett. **97**, 164301 (2006)]. Powerful underwater explosions and certain natural sources, such as underwater landslides, generate very low-frequency waves in water and air, for which both fluid buoyancy and compressibility simultaneously serve as restoring forces. In this paper, analysis of sound transmission through air-water interface is extended to acoustic-gravity waves (AGWs). It is found that, as for sound, the interface becomes anomalously transparent for sufficiently shallow compact sources of AGWs. Depending on the source type, the increase in wave power flux into air due to diffraction effects can reach several orders of magnitude. Physical mechanisms responsible for the anomalous transparency are discussed. Excitation of an interface wave by an underwater source is shown to be an important channel of AGW transmission into atmosphere, which has no counterpart in the case of sound.

10:30

5aPAb2. Sound control by temperature gradients. Jose Sanchez-Dehesa, Mitko I. Angelov, Francisco Cervera (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de vera s.n. ES-46022 Spain, jsdehesa@upvnet.upv.es), and Liang-Wu Cai (Kansas State Univ., Manhattan, KS 66506)

The acoustic mirage, like the optical mirage, is a consequence of temperature gradients existing in the air above Earth surface. Here, we report experiments showing that temperature gradients can be used to collimate an airborne ultrasonic beam. A very simple system consisting of two glass tubes heated by an internal resistive coil is used to create a temperature gradient producing the desired effect. Numerical simulations are also performed and results are in agreement with the experimental findings [Work supported by MICIIN of Spain.]

10:45

5aPAb3. Characterization of a lossy elastic plate thanks to the relationships between its reflection and transmission coefficients. Serge Derible and Alain Tinel (LOMC GOA, FRE CNRS 3102, Univ. Le Havre, 76600 Le Havre, France)

The normal reflection and transmission coefficients of a lossy elastic plate are measured and associated according to the outcomes of resonance scattering theory (RST) and S-matrix theory (SMT). The sum and the difference of those coefficients provide valuable tools, the transition terms, we use to characterize the plate. The absorption in the plate can be expressed via different manners recalled in the following. By means of a dimensionless loss factor, the velocities of the longitudinal and transverse waves of the studied aluminum plate are no more real but complex. The width of a resonance is the sum of two parts: the elastic width and the absorption one. In the vicinity of a resonance frequency, the transition terms obey the Breit–Wigner resonant form which makes it possible an easy determination of the two parts of the resonant width. There is a good experiment/theory agreement when an imaginary part is added to the Lamé coefficients of the aluminum plate. The loss factor is of the order 5.10–4. On the whole, the effect of the attenuation rises with frequency and reaches a maximal importance at the vicinity of a resonance frequency.

11:00

5aPAb4. Constraining the acoustic energy radiated from collisions of polypropylene balls. Joshua Riner and Andi Petculescu (Dept. of Phys., Univ. of Louisiana at Lafayette, P.O. Box 44210, Lafayette, LA 70504, riner@louisiana.edu)

Measurements of sound radiation from binary collisions of polypropylene balls were performed in order to study the collision dynamics and constrain the fraction of incident energy radiated as sound. In the experiments, one ball is released from directly above a stationary target ball. In the first experiment, the incident energy of the incident ball is varied by changing the drop height. The sound is recorded at an angle of 55 deg below the impact plane. Data from this experiment, fitted with a theoretical model, show a possible deviation from Hertzian behavior. The second experiment concens the radiation pattern of the collision process. The sound pressure was measured as a function of the polar angle θ. It is assumed that the problem is azimuthally symmetric, i.e., independent of the angle ϕ. The results are somewhat surprising: they show a dipole-like pattern that is asymmetric with respect to the impact line. The acoustic energy radiated during the impact, obtained by multiplying the collision time by the sound intensity integrated over a spherical surface centered at the impact point, is estimated as four orders of magnitude smaller than the incident energy. [The work was supported by the Louisiana Board of Regents.]

11:15

5aPAb5. Acoustical probe of the density fluctuations in a binary fluid mixture. Curtis F. Osterhoudt and Dipen N. Sinha (MPA-11, Los Alamos Natl. Lab., Los Alamos, NM 87545, cfo@lanl.gov)

Silicone oil was found to be fully miscible in "segregated" perfluorinated hydrocarbon (Novec fluid, by 3M) above a critical temperature slightly above room temperature and at atmospheric pressure. This constitutes a binary fluid mixture, and these have long been used as examples of continuous phase transitions, especially in regard to the phenomenon of critical opalescence [T. Andrews, Philos. Trans. R. Soc. London **159** (1869) and M. Smoluchowski, Ann. Phys. **21** (1906)]. During a phase transition in these mixtures, the fluids become spatially segregated, and the density fluctuation correlations are known to span a large number of length scales [L. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed., Part 1]. This behavior is reflected in massive fluctuations in sound speeds and viscosities in the mixtures near the critical point, and the resulting scattering behavior is probed by actively acoustically interrogating the mixture with techniques including swept-frequency acoustical interferometry and pulse-echo measurements. It is shown that with higher-power acoustical probes, the transition from unary to binary mixture may be prevented or enhanced. [This work was supported by Chevron USA.]

11:30

5aPAb6. Acoustic behavior of large encapsulated gas bubbles with resonance frequencies in the 50–100 Hz range. Mark S. Wochner, Kevin T. Hinojosa, Kevin M. Lee (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029), Theodore F. Argo, IV, Preston S. Wilson (The Univ. of Texas at Austin, Austin, TX 78712-0292), and Richard S. Mercier (Texas A&M Univ., College Station, TX 77843-3136)

A one-dimensional resonator technique was used to investigate the acoustic behavior of water containing large spherical and toroidal encapsulated air bubbles. Tethered balloons and inner tubes, respectively, were used to create the bubbles, which had effective spherical radii of approximately 5 cm. The number of balloons or inner tubes varied from 3 to 6, which resulted in void fractions from 1.5% to 5%. Effective mixture sound speeds for both shapes were inferred from the resonances of a water and bubble-filled waveguide 1.8 m in length. For the spherical bubbles, the Commander and Prosperetti (CP) model [J. Acoust. Soc. Am. **85**, 732–746 (1989)] quantitatively described the measured sound speed dispersion approaching the individual bubble resonance frequency and qualitatively predicted the frequency range of high attenuation, even in this regime of high-void fractions and discrete bubble distribution within the waveguide. Qualitative agreement was found between the CP model and the toroidal bubble measurements. Finite-element simulations of both experiments were performed, which help explain the discrepancy between the CP model and the measurements. [Work supported by Shell.]

11:45

5aPAb7. The characteristics of thunder on Titan. Andi Petculescu (Dept. of Phys., Univ. of Louisiana at Lafayette, P.O. Box 44210, Lafayette, LA 70504)

Even though *Huygens* has not detected acoustic signatures of Titanian lightning, electric discharge processes are thought to be very plausible in Titan's thick atmosphere dominated by nitrogen and methane. When correlated with electromagnetic signatures, thunder will very likely confirm the occurrence of lightning, both during a probe's descent and/or once it has landed. The presentation has two parts. In the first part, the mechanisms of Titanian thunder *generation* are discussed, in the context of the most up-to-date predictions of lightning on Titan, as well as relevant *Cassini-Huygens* data. In the second part, the *propagation* of thunder in Titan's troposphere is addressed. Here, the structure, composition, and ambient conditions of the satellite's atmosphere will be presented and accounted for. Current Titan lightning models predict cloud-to-ground discharge channels of ~20 km. Long-range Titanian thunder simulations are developed based on the far-field superposition of approximately 7000 *N*-waves arising from 3-m "elementary" sources. Two lightning geometries will be considered: linear and tortuous. (Tortuous channels are synthesized as a Gaussian distribution in the orientations of the 7000 elementary sources.) For illustrative purposes, to show the effects of atmospheric filtering, the thunder waveform described above will also be propagated in Venus' atmosphere.

FRIDAY MORNING, 23 APRIL 2010 DOVER A/B/C, 9:00 TO 11:30 A.M.

Session 5aPP

Psychological and Physiological Acoustics: Spectral, Temporal, and Complex Auditory Processing

Jennifer Lentz, Cochair
Indiana Univ., Dept. of Speech and Hearing Science, 200 S. Jordon Ave., Bloomington, IN 47405

Yi Shen, Cochair
Indiana Univ., Dept. of Speech and Hearing Science, 200 S. Jordon Ave., Bloomington, IN 47405

Contributed Papers

9:00

5aPP1. Effects of training in increment-detection and intensity-discrimination tasks. Walt Jesteadt, Harisadhan Patra, Melissa Krivohlavek, and Cynthia Rutledge (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Subjects use different cues when asked to detect a brief increment in a longer duration tone (increment detection) than when asked to discriminate between two tones of the same duration that differ only in level (intensity discrimination). To determine whether experience in one task would generalize or interfere with performance in the other, a group of six subjects completed 4000 trials in an increment detection task, followed by 800 trials of intensity discrimination, and final 800 trials of increment detection. A second group of subjects was tested in the reverse pattern, beginning with 4000 trials of intensity discrimination. The increment or signal was a 90-ms, 4-kHz tone. The pedestal or standard was a 490- or 90-ms, 4-kHz tone presented at 70-dB sound pressure level. A two-interval, forced choice adaptive procedure was used to estimate the level of the increment required for 71% correct. The results showed modest improvement over time for two subjects in increment detection and one in intensity discrimination, but at least one subject showed the opposite effect in each task. There was no generalization or interference between the two tasks and no correlation between increment detection and intensity discrimination across the 12 subjects. [Work supported by NIH.]

9:15

5aPP2. The relative contribution of adaptation and temporal integration in forward masking. Naveen N. Nagaraj and Jeffrey J. DiGiovanni (Auditory Psychophysics and Signal Processing Lab., School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701, nn328108@ohio.edu)

Forward masking has been studied extensively, but the mechanism underlying forward masking is still being debated. The aim of this study was to understand the relative contribution of adaptation and temporal integration on forward masking. In the first experiment, gap detection thresholds were measured for symmetrical and asymmetrical noise markers ranging from 10- to 30-dB SL in 5-dB increments. Further threshold measurements were made while keeping the level of the noise marker before the gap at a constant 30-dB SL while varying the level of noise marker after the gap between 10- and 25-dB SL. Noise marker sets with gap detection threshold of 20 ms or less were selected for the second experiment. In the second experiment, participants adjusted gap durations within an asymmetrical noise marker to be perceptually equal to that of symmetrical noise markers. The results will be discussed in the context of a temporal integration model, an adaptation model, and a hybrid model including both temporal integration and adaptation components.

9:30

5aPP3. Perceived continuity and pitch shifts. Christopher J. Plack (Human Commun. and Deafness Div, Univ. of Manchester, Manchester M13 9PL, United Kingdom, chris.plack@manchester.ac.uk) and Rebecca K. Watkinson (Univ. of Essex, Colchester CO4 3SQ, United Kingdom)

If a noise burst is introduced into the gap between two brief complex tone bursts, pitch discrimination improves to the level observed when the bursts really are continuous. In the experiment described here, it was shown that perceived continuity also affects sensitivity to envelope phase across the gap. Two complex tone bursts with unresolved harmonics and five regular envelope peaks ("pitch pulses") were presented sequentially, separated by a gap of two periods. When the relative phases of the two bursts were varied, such that the inter-pulse interval (IPI) across the gap was varied, the pitch of the whole sequence was little affected, consistent with previous results suggesting that the pitch integration window may be "reset" by a discontinuity. However, when the gap between the two bursts was filled with a noise with the same spectral envelope as the complex, variations in IPI had substantial effects on the pitch of the sequence. It is suggested that the presence of the noise causes the two tone bursts to appear continuous; hence resetting does not occur, and the pitch mechanism is sensitive to the phase discontinuity across the gap.

9:45

5aPP4. Frequency-specific measurements of the temporal modulation transfer function using narrowband noise. Yi Shen and Jennifer Lentz (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S Jordan Ave., Bloomington, IN 47405, shen2@indiana.edu)

Interpretation of frequency-specific measures of temporal acuity is often confounded by the coexistence of spectral and temporal cues. The current study presents a new experimental technique to measure the temporal modulation transfer function (TMTF), in which inherent envelope modulations in narrowband noise are used as the detection target. Modulation depths of the stimuli were manipulated using a signal processing algorithm that introduced no change in the stimuli's power spectra. In this way, spectral cues cannot cue the presence of modulation. TMTFs were measured using these narrowbands of noise [TMTF-NS (no spectral cues)] at four center frequencies (2, 4, 6, and 8 kHz) and bandwidths of 16, 24, 32, 48, 64, 96, and 128 Hz. TMTFs using low-pass-noise modulated puretones (TMTF-NBM) were also measured for comparison. In cases where the spectral cues are limited in TMTF-NBM (at low-modulation rates or at high-carrier frequencies), thresholds are very similar between the TMTF-NS and the TMTF-NBM, providing validation that this new technique measures temporal acuity similar to standard TMTF methods. In conditions where spectral cues are available in TMTF-NBM (high-modulation rates and lower carrier frequencies), higher thresholds are measured for TMTF-NS and a more reasonable estimate of temporal acuity is obtained.

10:00—10:15 Break

10:15

5aPP5. Another indication that information in "off-frequency" filters enhances sensitivity to envelope-based interaural temporal disparities. Leslie R. Bernstein and Constantine Trahiotis (Depts. of Neurosci. and Surgery (Otolaryngol.), Univ. of Connecticut Health Ctr., 263 Farmington Ave., Farmington, CT 06030, les@neuron.uchc.edu)

A recent study, [Bernstein, L. R., and Trahiotis, C. (**2009**). J. Acoust. Soc. Am. **125**, 3234-3242] employing "raised-sine" stimuli centered at 4 kHz revealed three outcomes concerning discriminability of envelope-based interaural temporal disparities (ITDs). First, graded increases in raised sine exponent led to graded decreases in envelope-based threshold-ITDs. Second, threshold-ITDs decreased with increases in modulation depth. Third, when the modulation depth was 25%, increasing the exponent led to especially large decreases in threshold-ITD. Overall, an interaural correlation-based model including stages mimicking peripheral auditory processing was shown to capture quite well most of those outcomes. Left unexplained, however, was the interaction between modulation depth and raised sine exponent. Here, we report new threshold-ITDs obtained while varying, factorially, depth of modulation (25% or 100%), raised-sine exponent (1.0 or 8.0), and modulation frequency (32, 64, 128, or 256 Hz). Those stimulus parameters were chosen specifically to shed light on the unexplained interaction. A quantitative theoretical analysis of the new data suggests that listeners can enhance their sensitivity to envelope-based ITDs by utilizing "off-frequency" filters centered above 4 kHz. [This research was supported by research grant NIH DC-04147 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.]

10:30

5aPP6. Masker modulation regularity and its effects on comodulation masking release. Emily Buss, Joseph W. Hall, and John H. Grose (UNC School of Medicine, Chapel Hill, NC 27599, ebuss@med.unc.edu)

Several authors have suggested that comodulation masking release is largest when performance in the narrowband masker condition is poor due to perceptual similarities between masker fluctuations and an added signal. The present study tested this hypothesis by manipulating the regularity of masker modulation between but not within listening intervals. The signal was a 500-Hz tone burst, and the masker was a continuous band of noise, either 25- or 1050-Hz wide. After filtering, the masker was amplitude modulated with one of three envelope patterns. The regular envelope was a smoothed 20-Hz square wave, composed of 10-ms raised-cosine ramps and 15-ms steady states. In the irregular envelope conditions, either the duration of the steady state or the modulation depth was randomly selected prior to each modulation period, with one exception: modulation was not randomly perturbed during listening intervals, when the signal might be presented. There were large individual differences, but masker envelope irregularity tended to elevate thresholds, particularly for the 25-Hz masker bandwidth. Results were consistent with more CMR under conditions for which there is greater informational masking in the narrowband masker condition.

10:45

5aPP7. Effects of modulated noise on detection of formant-like stimuli. Peggy B. Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455), Magdalena Wojtczak (Univ. of Minnesota, Minneapolis, MN 55455), and Yingjiu Nie (Univ. of Minnesota, Minneapolis, MN 55455)

One problem reported by listeners with hearing impairment (HI) might be that formant transitions in noise are not as salient as for listeners with normal hearing (NH). We hypothesize that a reduced ability to perceive formant transitions in noise could contribute to reduced masking release seen in HI listeners. In this study, we used pure-tone glides and harmonic complex simulations of formant transitions (shifts in intensity increments across frequency range of a formant transition over the transition duration) to test this hypothesis. Pure-tone glides were 30- or 60-ms long and were centered at 1250 or 2500 Hz, frequencies representing second formants in speech. They were presented in isolation or with unmodulated flanking tones (simulating F1 and F3) in quiet, steady noise, and gated noise. NH listeners were only slightly better than HI listeners at detecting the single gliding tone in quiet. Flanking tones adversely affected performance of the HI but not NH listeners. Performance of the NH listeners was somewhat better in gated than in steady noise, whereas the HI listeners performed similarly in steady and gated noise. The role of perception of formant transitions in masking release will be discussed. [Work supported by NIDCD R01-DC008306.]

11:00

5aPP8. The fluctuating-masker benefit for word identification with spectral or fine-structure distortions. Joshua G.W. Bernstein (Army Audiol. and Speech Ctr., Walter Reed Army Med. Ctr., Washington, DC 20307, joshua.g.bernstein@us.army.mil)

Normal-hearing (NH) listeners presented with speech processed to reduce spectral detail or remove temporal fine-structure (TFS) information show a reduced benefit from momentary dips in the level of a fluctuating masker. This has been interpreted as evidence that related deficits may underlie the reduced fluctuating-masker benefit (FMB) observed for the hearing impaired. However, the reduced FMB for processed stimuli could be attributable to the greater signal-to-noise ratio (SNR) required to yield performance levels equivalent to the unprocessed case for unmodulated noise (the reference condition). NH listeners were tested in the identification

of isolated words presented in stationary-noise, interfering-talker, and speech-modulated noise maskers. Different response-set sizes were used to offset performance differences between processed and unprocessed conditions. A validation experiment showed no effect of set size on the FMB when compared at a common stationary-noise SNR. A second experiment compared the FMB for unprocessed, spectrally smeared, and noise vocoded speech. When compared at the same SNR and percent-correct level (but different set sizes), processed and unprocessed stimuli yielded similar FMB. This suggests that for the conditions tested here, spectral or TFS distortions do not directly impair the ability to listen in the gaps of a fluctuating masker. [Sponsored by the Oticon Foundation.]

11:15

5aPP9. Psychometric properties of the coordinate response measure corpus with various types of background interference. David A. Eddins (Dept. of Otolaryngol., Univ. of Rochester, Rochester, NY 14642 and the Int. Ctr. for Hearing and Speech Res., Rochester Inst. of Technol., Rochester, NY 14623 USA.) and Chang Liu (Univ. of Texas at Austin, Austin, TX 78712)

The coordinate response measure (CRM) corpus has gained broad acceptance as a research tool for investigating speech intelligibility in background competition and has been widely used in studies of informational masking. Beneficial properties of the CRM include closed set format, limited linguistic content, and inclusion of target items that promote sound stream segregation based on the voice-identify of the target talker. Extensive research has demonstrated the psychometric characteristics of CRM target-word identification in the presence of one, two, or three competing talkers from the same corpus as well as long-term speech-shaped noise (LTSSN). Percent correct identification varies non-monotonically with signal-to-background ratio for one and two CRM-talker interference and is monotonic when the number of talkers exceeds 2 or is LTSSN. The CRM has even broader applicability when used with other types of background competition. The purpose of this study is to establish the psychometric characteristics of CRM target-word identification in various backgrounds with the goal of being able to determine when it is appropriate or not to use adaptive threshold procedures with the CRM corpus. Psychometric functions and thresholds based on adaptive tracking for identification of CRM target words presented in multi-talker babble and cafeteria noise will be discussed.

FRIDAY MORNING, 23 APRIL 2010 GRAND BALLROOM VII/VIII/IX, 8:00 A.M. TO 12:00 NOON

Session 5aSC

Speech Communication: Vowels, Phonation, and Tones (Poster Session)

Susan N. Nittrouer, Chair
Ohio State Univ. Medical School, Dept. of Otolaryngology, 915 Olentangy River Rd., Ste., 4000, Columbus, OH 43312

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

5aSC1. Phonetic learning *in utero*. Christine Moon (Dept. of Psych., Pacific Lutheran Univ., Tacoma, WA 98447, mooncm@plu.edu), Hugo Lagercrantz (Neonatology Unit, Karolinska Inst., Stockholm, Sweden), and Patrica K. Kuhl (Inst. for Learning & Brain Sci., Univ. of Washington, Seattle, WA)

The prevailing view is that newborn phonetic perception is *tabula rasa* because of poor transmission of the acoustic features of phonemes to the fetus. However, vowel information may be at least intermittently clear *in utero*. We tested 80 neonates (M = 32.8 h old, range 7–75) in the US and Sweden with English and Swedish vowels using an infant-controlled sucking procedure. Sucking activated 1 of 17 stimuli (a prototype and 16 variants) from the same vowel category, either the English /i/ or the Swedish /y/. Infants sampled through all 17 stimuli, presented randomly, one time. The dependent measure was mean number of sucks per stimulus. Results showed that the Foreign Vowel Group had significantly greater means to the prototype than the Native Group. A within-group analysis showed another Foreign-Native Group difference. Infants in the Foreign Group had significantly more sucks to the prototype compared to the variants, whereas the Native Group treated the prototype and variants equivalently. These results require us to re-evaluate assumptions about availability of speech sounds to the fetus and the state of speech perception at birth.

5aSC2. Longitudinal analysis of vowels in infant-directed speech. Kerry E. McColgan, Nan Bernstein Ratner, and Rochelle Newman (Dept. of Hearing and Speech Sci., Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742, kmccolgan@hesp.umd.edu)

Adult-directed conversational speech (ADS) is characterized by a high level of acoustic imprecision which should theoretically cause problems for infant language learning. Whether infant-directed speech (IDS) is clearer than ADS is disputed. However, differences in prosody [Fernald (1989)], utterance length [Cooper (1997)], and acoustic features [Bernstein Ratner (1984) and Malsheen (1980)] have been found. Such differences have the potential to assist children in decoding language input [Fernald (1985)]. In prior research, the child's age and language ability appear to affect speech clarification [Bernstein Ratner (1996)]. This study examines whether vowel clarification is a function of the infants' age. More than 75 mother-infant dyads are being followed from 7 months to 2 years of age. Data reported here are from 10 dyads whose data have been fully analyzed. Plots of the first and second formant values for vowels in words spoken in IDS and ADS were compared at 7 and 11 months of age. Vowels to 7-month-old infants appear to show less clarification than those to 11-month infants, although individual patterns of clarification are evident. This study analyzes whether such variability in vowel clarification affects the rate of early infant vocalizations prior to language production. [Work supported by NSF BCS074512.]

5aSC3. Vowel space of 3- and 4-year-old children. Rebecca W. McGowan, Margaret Denny, Richard S. McGowan (CReSS LLC, 1 Seaborn Pl. Lexington, MA 02420), and Susan Nittrouer (The Ohio State Univ., Columbus, OH 43212)

Children learn to produce speech within the constraints of vocal tracts that are smaller and configured differently from those of adult speakers. To study children's speech development in relation to their vocal tract growth, the acoustic characteristics of vowels spoken by ten children, recorded at 6 month intervals between the ages of 36 and 48 months and a subset between 18 and 48 months, are analyzed. The first three formant frequencies of each token are measured and tokens are categorized by intended vowel. Formant frequencies are compared both among vowel categories at each age and

within categories over time to give an understanding of the development of children's vowel space. Changes in children's ability to differentiate among vowels in F1-F2 space over time are investigated. The development of vowel space is considered in the context of growth of and changes in the proportions of the vocal tract and improved control of the articulators. [Work supported by NIDCD-0001247 to CReSS LLC.]

5aSC4. Neural coding of formant-exaggerated speech in infants and adults. Yang Zhang, Sharon Miller, Tess Koerner, and Edward Carney (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillbury Dr. SE, Minneapolis, MN 55455)

Speech scientists have long proposed that formant-exaggerated speech plays an important role in phonetic learning and language acquisition. However, there have been very little neurophysiological data on how the infant brain and adult brain respond to formant exaggeration in speech. We employed event-related potentials (ERPs) to investigate neural coding of formant-exaggerated speech sounds. Two synthetic /i/ vowels were modeled after infant-directed speech data and presented in alternating blocks to test the effects of formant exaggeration. The fundamental frequencies of the two sounds were kept identical to avoid interference from exaggerated pitch level and range. For adult subjects, non-speech homologs were also created by using the center frequencies of the formants to additionally test whether the effects were speech-specific. In the infants (6 to 12-month olds), ERP waveforms showed significantly enhanced N250 and sustaining negativity responses for processing formant-exaggerated speech. In adults, enhancement was observed in the N100 component for the speech stimuli but not the homologous non-speech sounds. Collectively, these results provide the first evidence that formant expansion in infant-directed speech enhances neural activities for phonetic encoding, which may facilitate phonetic learning and language acquisition regardless of the age factor [Zhang *et al.* (2009). Neuroimage 46, 226–240].

5aSC5. Development of vowel production in young children with cochlear implants. Kyoko Nagao, Allegra Cornaglia, and H. Timothy Bunnell (Ctr. for Pediatric Auditory and Speech Sci., Nemours A.I. duPont Hospital for Children, Wilmington, DE 19803, nagao@asel.udel.edu)

Past research on children with cochlear implants (CIs) focused on improvements in speech perception, but few speech production studies have been reported. The literature on speech development in prelingual children with CIs is particularly sparse. This is partially due to difficulty transcribing young children's productions. The aim of this study was to examine whether developmental changes in vowels can be captured without phonetically transcribed data. The subject, a prelingually deaf child 22 months old at the start of the study, is one of six children with CIs participating in an ongoing longitudinal study. Audio and video data were collected once or twice per week during auditory/verbal therapy at the Alfred I. duPont Hospital for Children. One author perceptually categorized the child's vowel-like vocalizations in terms of broad height (HIGH, MID, and LOW) and place (FRONT, MID, and BACK) features. Acoustic features (barkscale cepstral coefficients) were then used to train linear discriminant functions on the same vowel data and used to predict perceptual vowel categories. In preliminary results 62% of the tokens were correctly classified for place and 67% for height. Results on generalizability of the features will be presented. [Work supported by NIDRR Grant H133E080003 to the RERC on Hearing Enhancement.]

5aSC6. Sensorimotor maps and vowel development in English, Greek, and Korean: A cross-linguistic perceptual categorization study. Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., Minneapolis, MN 55455, munso005@umn.edu), Lucie Mnard (Universit du Quebec, Montreal, QC 3C 3P8, Canada), Mary E. Beckman (Ohio State Univ., Columbus, OH 43210), Jan Edwards, and Hyunju Chung (Univ. of Wisconsin, Madison, WI 53705)

Learning to speak involves learning the association between different articulatory maneuvers and their associated auditory-perceptual characteristics. These associations may change over the course of development as a consequence of the nonlinearities inherent in vocal-tract growth. A set of recent studies by Ménard and colleagues has examined these developmental changes by studying adults' categorizations of synthesized vowels based on an articulatory model of the growing vocal tract. These have shown that vowels modeled on younger vocal tracts tend to be perceived as more front than those modeled on older vocal tracts. They have also shown language-specificity in the vowels that are identified in growing vocal tracts, with French listeners labeling fewer sounds as /u/ than English ones, presumably because the high-vowel space of French includes three vowels /i y u/, while English has only two. The current experiment expands on this by examining the perception of synthesized vowels based on seven vocal-tract growth stages by speakers of three languages with very different vowel systems, American English, Korean (seven vowels), and modern Greek (five vowels). Preliminary analyses show language-specific patterns in the perception of vowels in developing vocal tracts. [work supported by NIDCD 02932 and NSF grants BCS0729140 and BCS0729277]

5aSC7. Korean listeners' sensitivity to language-specific phonetic details of children and adults' vowel production in five different languages. Hyunju Chung, Jan Edwards, Gary Weismer, Eunjong Kong (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705, hchung23@wisc.edu), Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455), and Mary E. Beckman (Ohio State Univ., Columbus, OH 43210-1298)

Systematic differences in vowel production across languages have been demonstrated, even for so-called "shared" vowels. Starting from 10 months of age, language-specific vowel production patterns emerge in babbling [Boysson-Bardies *et al.* (1989)]. Vowel production data from a word repetition task in our laboratory also showed language-specific patterns emerging at 2 years of age. While there is research on adult and infant perception of non-native vowels produced by adults, there is little work on adult perception of children's non-native vowels of different languages. The current study examined Korean adult listeners' perception of vowels from five languages, Cantonese, English, Greek, Japanese, and Korean, produced by 2-, 5-year-old children, and adults. Consonant-vowel and vowel only stimuli were extracted from our vowel production data. Target vowels included /i/, /a/, and /u/, following velar and alveolar stops. Twenty native Korean adult listeners were asked to identify the vowels and rated their goodness. Results will be analyzed using both the formant frequencies and listeners' judgments using a mixed effect model. [Supported by a Fulbright fellowship, NIDCD Grant 02932, and NSF Grant 0729140.]

5aSC8. Coordination of the first and second formants of the Mandarin triphthong /iau/ revealed by adaptation to auditory perturbations. Shanqing Cai, Satrajit S. Ghosh, Frank H. Guenther, and Joseph S. Perkell (Speech Commun. Group, Res. Lab. of Electrons, MIT, Cambridge, MA 02139)

Investigations of auditory feedback control of supra-glottal articulatory movements traditionally have focused on quasi-static speech sounds like monophthongs. Only recently has attention been directed toward time-varying sounds. To study the role of auditory feedback in the formation of formant trajectories in time-varying vowels, we perturbed the second formant frequency (F2) trajectory as 20 native speakers produce the Mandarin triphthong /iau/ and measured their patterns of auditory-motor adaptation. The subjects changed F2 in directions opposite to the perturbations, not unlike adaptive responses seen in previous studies. Surprisingly, concurrent with the F2 compensations, the subjects significantly altered the first formant (F1), the unperturbed formant, in ways that assisted cancellation of the perturbations to the location and orientation of the formant trajectory in the F1×F2 space. Care was taken to rule out that the F1 changes were byproducts of the F2 compensations. These observations indicate that F1 and F2 are planned in a coordinated fashion for the triphthong /iau/, a property of the speech system not yet revealed through studying quasi-static sounds. The implication of this F1-F2 synergy for the spatiotemporal nature of vowel auditory targets will be discussed. The patterns of generalization to a set of unperturbed vowels will also be presented.

5aSC9. Position and height asymmetries in hiatus resolution: Are they phonetically driven phenomena? Hijo Kang (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376, hikang@ic.sunysb.edu)

Hiatus is not preferred in many languages and is resolved in various ways. Two typological patterns have been found in hiatus resolution. When hiatus is resolved by a weakening (deletion or gliding) of one vowel, (1) V1 is more likely to be the target than V2 [Casali (1996)] and (2) high vowels are more likely to be the target than non-high vowels [Rosenthall (1997)]. If

these patterns are due to human articulatory and/or auditory mechanisms [Ohala (1993)], the patterns would be expected to reflect variations in ordinary speech. In this study Korean CV1V2 nonce words produced by six Korean speakers are analyzed to test whether V1 and high vowels in a realized hiatus are actually weakened in terms of duration. The results showed that V1 is shorter than V2 irrespective of speech rate [$F(1,5)=12.90$, $P<0.02$], which was not found in CVCV or CVGV words. It was also found that high vowels are shortened in fast speech to a more degree than non-high vowels [$F(1,5)=7.16$, $P<0.05$]. The formant analysis showed that vowel-to-vowel coarticulation effects are greater in CV1V2 than in CVCV, which means that vowels in hiatus are more vulnerable to misperception and consequently, sound change.

5aSC10. Lexical frequency and Japanese vowel devoicing. Kuniko Nielsen (Dept. Linguist., Oakland Univ., 320 O'Dowd Hall, Rochester, MI 48309, nielsen@oakland.edu)

The current study aims to examine the relationship between Japanese vowel devoicing and lexical frequency in light of the probabilistic reduction hypothesis [Jurafsky *et al.* (2001)], which predicts that more probable words are (phonetically) reduced more than less probable words. In Tokyo Japanese, high vowels become devoiced when they occur between two voiceless consonants. Similar devoicing processes occur in many languages and are generally considered to be part of the vowel reduction processes [e.g., Dauer (1980); Kohler (1990)]. Although Japanese vowel devoicing has been studied extensively from both phonetic and phonological viewpoints, its lexical property, namely, the effect of lexical frequency, is as of yet unknown. If Japanese devoicing is a phonetic reduction process, we would expect to see a stronger degree of devoicing among words with higher lexical frequency. Twenty-four speakers of Tokyo Japanese recorded 60 words which include various devoicing environments. Contrary to the prediction by the probabilistic reduction hypothesis, our preliminary results revealed a significantly higher rate of devoicing [$F(1,22)=24.89$, $p<0.01$] as well as shorter vowel duration [$F(1,22)=4.61$, $p<0.05$] among low-frequency words than high-frequency words. The implications of these results on the nature of Japanese vowel devoicing will be discussed.

5aSC11. Comparison of the formant frequencies F3 and F4 on a three-dimensional vowel chart. Toshio Isei-Jaakkola (Dept. of English Lang. & Culture, Chubu Univ., 1200 Matsumoto, Kasukagi, Aichi 487-8501, Japan, tiseij@isc.chubu.ac.jp), Takatoshi Naka (Chukyo Univ., Toyoda, Aichi 470-0393, Japan), and Keikichi Hirose (Univ. of Tokyo, Bunkyo, Tokyo 113-8656, Japan)

We have developed a three-dimensional vowel chart that should fit into a quadrilateral IPA vowel chart better and later we have added more visually effective functions to it. On the chart we can see not only the locations of F1 and F2 with two scales of *x* and *y*, where *x* is F1 and *y* F2, but also F1, F2, and F3 as well, with three scales of *x*, *y*, and *z*, where *x* is F1, *y* F2, and *z* F3. Thus, this enables us to compare the vowel locations of any language (multilingual vowel presentations), vowels uttered by male with those of female (thus, gender difference), vowels of adults with those of children (thus, age difference), foreign language learners with target vowels, or vowels uttered by the same language learners. This time, we focus more on the roles of the respective formants, particularly on the roles of F3 and F4. It has been clarified that F3 is more related to lip-spreading. We want to show that F4 may be more related to lip-protrusion. For this purpose, we have further added a selection of F3 and F4 functions to the chart. [Work supported by JSPS and Chubu University Grant (A).]

5aSC12. Psychometric functions of vowel detection and vowel identification in multi-talker babble. Maggie Miller, Jessica Walker, Kayla O'Brien, and Chang Liu (Dept. of Commun. Sci. and Disorder, The Univ. of Texas at Austin, 1 Univ. Station A1100, Austin, TX 78712)

Psychometric functions of vowel detection and vowel identification were measured in multi-talker babble for young normal-hearing listeners. A four-interval forced-choice procedure was used to examine the accuracy of vowel detection in babble with speech level presented from 0- to +15-dB sensation level relative to vowel detection thresholds obtained with method of limits. The accuracy of vowel detection was significantly influenced by vowel category and sensation level. The threshold of vowel detection for each vowel and each listener was defined as the speech level at which 71% accuracy of vowel detection was reached. Vowel identification was then measured in babble with vowel levels presented from 0- to 12-dB sensation level relative to individual thresholds of vowel detection, using a close-set 12-choice procedure. Results suggest that vowel identification was significantly affected by vowel category and sensation level. Altogether, the results of vowel detection and vowel identification indicate that, given the same signal-to-noise ratio, vowels are not equally audible and identifiable, possibly due to the fact that some vowels are more audible than others, and that the slope of psychometric functions of vowel identification is vowel-dependent.

5aSC13. Vowel-inherent spectral change in isolated vowels and consonant vowel consonants. Peter F. Assmann (School of Behavioral and Brain Sci., The Univ. of Texas at Dallas, Richardson, TX 75083), Terrance M. Nearey (Univ. of Alberta, Edmonton, AB T6G 2E7, Canada), and Michael Kiefte (Dalhousie Univ., Halifax, NS B3H 1R2, Canada)

To study the interaction of vowel inherent spectral change (VISC) and consonant context on vowel formant patterns, we recorded 15 vowels /i ɪ eɪ ɛ æ ʌ ɑ ɔ oʊ ʊ u ɝ ɑɪ ɑʊ ɔɪ/ in each of 14 consonant environments (pVp, pVb, bVb, bVp, tVt, tVd, dVt, dVd, kVk, kVg, gVk, gVg, hVd, and V) spoken by 10 men and 10 women from the north Texas region. Analysis of vowel formant frequency trajectories confirmed the reliable presence of VISC across talkers for a range of consonant environments. Vowels showing stable patterns of VISC included those acknowledged as diphthongs in most North American English dialects, /ɑɪ/, /ɑʊ/, /ɔɪ/, /eɪ/, /oʊ/, as well as /ɪ/, with more variable movement patterns for /ɛ/, /æ/, and /u/. Of particular interest were cases where the formants F1 and/or F2 showed a "switchback" pattern of movement, with initial movement in the expected direction for the vowel (as observed in isolated vowels) and subsequently toward the"locus" for the final consonant. Parallel analyses of VISC in three dialects (north Texas, Nova Scotia, and Alberta) are currently underway and will be reported at the meeting along with statistical modeling results [Nearey (this meeting)].

5aSC14. Assessing acoustic measures of the spontaneous phonetic imitation of vowels. Molly E. Babel (Dept. of Linguist., Univ. of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada, mbabel@interchange.ubc.ca) and Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455)

It is well established that people imitate fine phonetic detail of another talker in shadowing tasks [Goldinger, Psych. Rev. **105**, 251–279 (1998)] and in interactive conversation [Pardo, J. Acoust. Soc. Am. **119**, 2382–2393 (2006)]. That is, the acoustic characteristics of a model talker's production (MTP) are more similar to a participant's shadowed production (PSP) than they are to that participant's baseline production (PBP), typically elicited in a reading task. This presentation compares two methods of assessing the acoustic distance between the vowels in PSP and PBP monosyllabic words taken from two previous studies [Babel, thesis, University of California (2009); Kaiser & Munson, (unpublished)]. One of these measures is the F1/F2 Euclidean distance between PSP and PBP vowels. This measure does not take into account the direction of the difference. The other [adapted from Titze, *Principles of Voice Production* (1994)] compares the distance and direction of the F1 and F2 values in the PSP and PBP vowels relative to those of the MTP. The merits of each of these methods are assessed by comparing them to measures of listener judgments of the similarity of PSPs and PBPs to the MTPs in AXB perception tasks.

5aSC15. The relationship between fundamental frequency and vowel quality. Santiago Barreda and Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, 4-32 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada, sbarreda@ualberta.ca)

There is disagreement over the role fundamental frequency (f0) plays in the determination of vowel quality. Some claim [e.g., D. Smith *et al.*, J. Acoust. Soc. Am. **117**, 305–318 (2005)] that changes in f0 do not affect vowel quality at all. Others claim there is relationship between f0 and vowel quality, whether direct or indirect [see T. M. Nearey and P. F. Assmann, in *Experimental Approaches to Phonology*, edited by M. J. Solé and P. S. Beddor (Oxford University Press, Oxford, 2007), pp. 246–269 for a review]. To test these theories, a series of experiments was carried out in which participants were asked to make simultaneous speaker and vowel quality judgments. Participants were presented with a synthetic vowel continuum

spanning from [æ] to [ʌ] matched with several different f0s and higher formants. Participants were asked to rate vowel quality on a continuous scale ranging from completely [æ] to completely [ʌ] and to indicate the size and gender of the speaker. Results indicate that there is a complicated, indirect relationship between f0 and vowel quality, and that shifts in vowel quality caused by changes in f0 are a result of changes in assumed speaker properties.

5aSC16. Improved representation of variance in measures of vowel merger. Lauren Hall-Lew (Phonet. Lab., 41 Wellington Square, Oxford OX1 2JF, UK)

Current measures of vowel merger, such as the Euclidean distance between averages, have only been able to capture some of the variability between two given vowel clusters. Reliance on averages obscures the amount of variability within a given vowel class, while other techniques, such as calculating distance between minimal pairs, rely on few tokens per speaker. Hay *et al.* (2006) introduced an alternative approach that accounts for the variability between two vowel clusters, taking formant values as input, rather than averages. The measure is the Pillai-Bartlett statistic [Baayen (2008)], an output of a multivariate analysis of variance (MANOVA), which represents the proportion of one variance that can be predicted by another variance. A higher Pillai value indicates a lower degree of overlap between two vowel clusters in F1/F2 space. Since the value is derived from a MANOVA, Pillais can account for known internal factors influencing the production of merger, such as phonological environment, thereby reducing the need to obtain minimal pair lists. This talk argues for using Pillais as measures of merger by comparing results from low back merger in California English [Hall-Lew (2009)] with the analysis of front vowel merger by Hay *et al.* (2006).

5aSC17. An acoustic study of front rounded vowels in Shetland dialect. Man Gao and Peter Sundkvist (Acad. of Humanities and Media Studies, Dalarna Univ., SE-79160 Falun, Sweden)

This paper presents an acoustic analysis of front rounded vowels (FRVs) in the dialect spoken in the Shetland Islands, the northernmost locality of the British Isles. FRVs are typologically marked and estimated to occur in only 6.6% of the world's languages [I. Maddieson, in Haspelmath *et al. The World Atlas of Language Structures* (2005)]. Their occurrence in the Shetland dialect is, at least partly, attributable to a Scandinavian substratum language. There is significant variation across the archipelago regarding several aspects such as (1) the number of lexically contrastive FRVs, (2) phonetic quality (close to half-close), (3) contrastive length, and (4) lexical distribution and support. This paper presents an investigation of three speakers from one locality in which FRVs have retained firm lexical support. The issues addressed concern the dialect's overall acoustic vowel space (based on F1, F2, and F3), the position of FRVs within the acoustic space, and what the contrasts among FRVs and other adjacent vowels appear to rest on acoustically. Special focus is directed to phonetic contexts that support the greatest number of vowel contrasts and display the most crowded acoustic vowels spaces.

5aSC18. A new non-linear regression model for formant trajectories in English monosyllables incorporating dual targets for vowels. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, 4-32 Assiniboia Hall, Edmonton, AB, T6G 2E7, Canada, t.nearey@ualberta.ca)

A new non-linear regression model is proposed to characterize the formant trajectories of the vocalic portion of English CVC syllables in the data described by Hillenbrand *et al.* [J. Acoust. Soc. Am. **109**, 748–763 (2001)]. The modeling framework builds on work of Broad and Clermont [J. Acoust. Soc. Ame. **81**, 155–165 (1987)], wherein formant trajectories were modeled via three additive components: (1) a single vowel target, (2) an exponential approach (in time from onset) toward the vowel target from an initial consonant onset value, and (3) an exponential approach (in time from offset) toward the vowel target from a final consonant offset value. The new model extends this to allow for a dual specification (nucleus + offglide) of the vowel targets [T. Nearey and P. Assmann, J. Acoust. Soc. Am. **80**, 1297–1308 (1986)]. Initial results suggests that inclusion of a second vowel target provides substantial reduction (about 40%, 23% and 7% respectively for F1, F2, and F3) of error variance on trajectories averaged across 12 speakers. More detailed statistical analyses of variants of the new model are underway and will be reported for both the data described above and that reported on by Assmann *et al.* [this meeting].

5aSC19. The effect of increased loudness on anticipatory coarticulation of steady-state formants in a vowel-consonant-vowel context for patients with Parkinson's disease. Jennifer A. Kamphaus, Joan E. Sussman, Elaine T. Stathopoulos, Jessica E. Huber, Kelly C. Richardson, and Brit A. Boyarsky (Dept. of Communicative Disord. and Sci., Univ. at Buffalo, SUNY, 3435 Main St., Buffalo, NY 14214)

The current investigation examines the effect of increased loudness on anticipatory coarticulation of the vowels [i, ɑ, u] on the preceding schwa vowel for patients with Parkinson's disease. Four speakers, three males and one female with speech severity levels between 8% and 49%, are involved in pre-treatment, treatment, and post-treatment conditions. During treatment, participants use an auditory device which plays speech-babble into one ear while they are speaking. The speech-babble competition results in an automatic loudness increase (Lombard effect) in the talker wearing the device. Participants produced target nonsense CVC words embedded in carrier sentences at comfortable loudness levels and at an average of 3 dB above that level. The steady-state F1, F2, and F3 formant frequencies of the preceding [ə] were analyzed to determine whether increased loudness results in increased anticipatory coarticulation from the [i, ɑ, u] vowels in the CVC words, as treatment progresses. Results using steady-state formant frequency measures from this study will be compared with previous studies that used F2 trajectory measures [Tjaden and Sussman (2006); Weismer *et al.* (1995); Tjaden (2003)]. (Work supported by NIH 1R01DC009409-01)

5aSC20. Investigating within and between talker variability in the optical characteristics of American English vowels. Edward T. Auer, Jr. and Laura Welch (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, auer@ku.edu)

Previous research investigating vowel production provides evidence of substantial within and cross-talker variation in acoustic characteristics. The current study is designed to investigate whether similar within and cross-talker variation exists in the optical characteristics of American English vowels. Additionally, potential shared acoustic-optical variation across tokens will also be examined. Measures taken from three-dimensional motion data for a set of 13 markers positioned around the lips, cheeks, and chin sampled at 100 frames/s with millimeter spatial resolution along with simultaneous video (50 frames/s) and audio for each of ten repetitions of 11 vowels spoken by multiple talkers will be reported. Preliminary analysis of motion and acoustic data from12 repetitions of 11 vowels spoken by one male talker provides evidence of substantial within vowel variation in both vertical and horizontal lip separations measured at the midpoint of the vowel as determined by analysis of the acoustic signal. Correlations were observed between horizontal lip separation and second formant frequency ($r = 0.48$) and third formant frequency ($r = 0.66$) at the vowel midpoint. No correlations were obtained with vertical lip separation measure. Results will be discussed in terms of potential implications for talker intelligibility in visual only presentation conditions.

5aSC21. Effects of ambient pressure and gas mixture on a numerical model of subglottal acoustics and vowel spectra. Steven M. Lulich (Dept. of Psych., Washington Univ., 1 Brookings Dr., Saint Louis, MO 63130, slulich@artsci.wustl.edu)

Divers' speech in hyperbaric heliox environments becomes significantly distorted, making communication between divers and with surface-based personnel difficult. Previous research identified two basic mechanisms underlying this distortion. First, the speed of sound increases significantly in heliox, causing a linear increase in the second and higher formant frequencies. Second, in hyperbaric conditions the specific impedance of air in the vocal tract approaches that of the vocal tract walls, leading to a nonlinear increase in the first formant frequency. Effects of ambient pressure and gas mixture on subglottal acoustics and subglottal-vocal tract coupling in vowel spectra have not been sufficiently investigated. In this paper, such an investigation is carried out using a numerical model of subglottal and vowel acoustics. It is shown that subglottal resonance frequencies increase linearly as the speed of sound increases, and the corresponding subglottal pole-zero pairs in vowel spectra become disproportionately prominent as

ambient pressure increases. Increased acoustic coupling between the subglottal system and the vocal tract in hyperbaric heliox therefore constitutes a third significant source of speech distortion. The results have implications for understanding subglottal coupling in vowels and for heliox speech unscrambling techniques. [This work was supported in part by NSF Grant No. 0905250.]

5aSC22. Developing vowel mappings for an interactive voice synthesis system controlled by hand motions. Karl I. Nordstrom (Media and Graphics Interdisciplinary Ctr., Univ. of British Columbia, Forest Sci. Ctr. Bldg., FSC 3640-2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada, karl@karlnordstrom.ca), Sidney Fels, Cameron D. Hassall, and Bob Pritchard (Univ. of British Columbia, Vancouver, BC, V6T 1Z2, Canada)

This study investigates vowel mappings for a voice synthesizer controlled by hand gestures for artistic performance. The vowel targets are on a horizontal plane navigated by the movement of the right hand in front of the performer. Two vowel mappings were explored. In one mapping, the vowels were evenly distributed in a circle to make the vowel targets easier for the performer to find. In the other mapping, the vowels were arranged according to the F2 versus F1 space. Linear hand motions were then made through the vowel space while plotting the formant trajectories. The evenly distributed mapping resulted in formant trajectories that were not monotonic; the F1 and F2 pitch contours varied up and down as the hand carried out the linear motions. This had the unintended result of producing multiple diphthongs. In contrast, the F2 versus F1 mapping enabled the performer to create monotonic formant trajectories and the perception of a single diphthong. The performer found it easier to speak and sing through the system when a single linear hand motion resulted in a single diphthong. [This project was supported by Canada Council for the Arts, Natural Sciences and Engineering Council of Canada, and Media and Graphics Interdisciplinary Centre.]

5aSC23. Psychometric functions for rough voice quality. David A. Eddins (Univ. of Rochester, 2365 S. Clinton Ave., Ste. 200, Rochester, NY 14618, david_eddins@urmc.rochester.edu) and Rahul Shrivastav (Univ. of Florida, Gainesville, FL 32611)

Dysphonic voices arising from laryngeal lesions are generally believed to vary across three perceptual dimensions—breathiness, roughness, and strain. This experiment sought to evaluate the psychometric functions for roughness and to determine the acoustic cues for its perception. Ten normal-hearing listeners, all native speakers of American English, were recruited for this experiment. The perception of roughness in dysphonic voices was evaluated for 34 voices from the Kay Elemetrics Disordered Voice Database (Kay-Pentax, Inc., Lincoln, NJ). Listeners compared the roughness of these test stimuli against that for a sawtooth wave with a 40-Hz squared-cosine amplitude modulation (AM). The modulation depth of the sawtooth wave was varied from low to high in a paired comparison task. Psychometric functions for roughness were derived by plotting the listener response for each stimulus pair against the modulation depth of AM sawtooth wave. The data for each stimulus were fitted with a logistic function, and various parameters (threshold, slope, and saturation point) were computed. These parameters were then compared to various acoustic measures obtained from each voice to determine candidate acoustic cues for roughness.

5aSC24. Glottal aperture monitoring with external lighting and sensing photoglottography. Kiyoshi Honda (LPP, UMR7018 CNRS-Univ. Paris3, 19 rue des Bernardins, 75005 Paris, France), Shinji Maeda (TELECOM ParisTech, F-75634 Paris Cedex 13, France), and Tatsuya Kitamura (Konan Univ., Kobe 658-8501, Japan)

A non-invasive photoglottographic method has been developed for monitoring glottal aperture changes during speech for the purpose of phonetic and clinical studies. The system includes light-source and sensor units both placed externally on the neck. An LED light source on the side of the neck illuminates the hypopharynx diffusely, and a photo-sensor unit on the front neck below the cricoid cartilage detects light passed through the glottis. An ambient light rejection circuit was newly added to avoid the effect of room light. The photoglottography (ePGG) system is free from interference due to tongue retraction and thus operational both in high- and low-vowel environments, while it is susceptible to articulatory movements of the jaw and larynx. We will present new ePGG/airflow data to explain why the apparent word-initial strengthening of glottal opening occurs in our ePGG, as often observed in other previous PGG data. [Work supported by ESAPVI/ANR and Kakenhi 21300071.]

5aSC25. Breathy phonation in Gujarati: An acoustic and electroglottographic study. Sameer ud Dowla Khan (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853, sameeruddowlakhan@gmail.com)

Recent electroglottographic research on Gujarati [Khan (2009)] reveals that speakers consistently distinguish breathy and modal vowels (e.g., /bar/ "twelve" vs /bAr/ "outside;" capitalization represents breathiness) by both closed quotient and closing velocity. Despite this interspeaker uniformity measured at the voice source, the acoustic correlates of the contrast measured in the speech output vary considerably across speakers. While some speakers mark breathiness with a larger H1-H2 difference, others use lower periodicity (CPP) or greater changes in intensity (rms energy). It is unclear how breathy vowels are contrasted with modal vowels following breathy consonants (e.g., /bAr/ "outside" vs /Bar/ "burden") or to what extent the preceding consonant manner can affect the spectral acoustics of the vowel (e.g., /pAr/ "mountain" vs /vAn/ "vehicle"). To better capture the variation in both vowel and consonant phonations, the current study examines acoustic and electroglottographic data collected from naturalistic productions of near-minimal sets from a larger pool of Gujarati speakers, using a wider range of vowel and consonant types. Preliminary results suggest that while all speakers distinguish phonation types in both consonants and vowels based on electroglottographic measures, the acoustic cues in the output are far more complex. [Work supported by NSF.]

5aSC26. Effects of asymmetric vocal fold activation on phonation. Dinesh K. Chhetri and Juergen Neubauer (Div. of Head and Neck Surgery, 62-132 CHS, 10833 Le Conte Ave., Los Angeles, CA 90095, dchhetri@mednet.ucla.edu)

The effects of asymmetric vocal fold activation on phonation were studied using an *in vivo* canine model and graded neuromuscular stimulation of the recurrent laryngeal nerve (RLN). One RLN was stimulated at 21 levels of graded stimulation, from threshold activity to maximal contraction, at three levels of contralateral vocal fold contraction (low, medium, and high). Phonation onset pressure (Pth) was recorded at each condition. The effects of airflow and cricothyroid (CT) muscle activation were studied by varying airflow and CT activation levels. Pth remained relatively constant over nearly all conditions while the level of graded stimulation needed to reach Pth varied. Increasing airflow and higher level of contralateral vocal fold activation allowed Pth to be reached at lower levels of graded stimulation, while activation of CT muscles had the opposite effect. At maximal CT activation, Pth was reached only at high levels of contralateral vocal fold activation but Pth was doubled. Results show that in vocal fold paresis /paralysis, phonation onset can be achieved by increasing either airflow or activation level of the normal vocal fold, and that CT activation increases the phonatory effort. These results provide further insights into the compensatory mechanisms required for phonation in these common laryngeal pathologies.

5aSC27. An acoustic analysis of pulmonic ingressive speech in the Shetland Islands. Peter Sundkvist, Man Gao (Acad. of Humanities and Media Studies, Dalarna Univ., SE-79160 Falun, Sweden, psn@du.se), and Gunnel Melchers (Stockholm Univ., SE-10691 Stockholm, Sweden)

The use of a pulmonic ingressive airstream mechanism in the pronunciation of certain discourse particles, typically variants of "yes" and "no," is a well-known and salient feature of Scandinavian languages. It has been suggested, however, that this may be a more general North Atlantic phenomenon—occurring as far west as Newfoundland and New England—which spread via migration and trade routes. Unfortunately, there seem to be very few audio recordings available from areas other than Scandinavia and Newfoundland (perhaps partly attributable to various elicitation difficulties) and very little acoustic analysis has been presented [E. Thom, MA thesis, UCL (2005); R. Eklund, J. Int. Phonetic Assoc. **38**, 235–325 (2008)]. This paper contains a study of ingressive discourse particles in the Shetland Islands, which have strong historical links to Scandinavia. A significant number of ingressives were found in field recordings from 1980–1982. This paper presents an acoustic pilot study of ingressive dis-

course particles, focusing on issues such as formant structure and voicing /phonation characteristics. A comparison is also made with previous analyses [E. Thom, MA thesis, UCL (2005); R. Eklund, J. Int. Phonetic Assoc. **38**, 235–325 (2008)] as well as audio samples from other localities within the North Atlantic region.

5aSC28. The surface wave model of phonation threshold pressure and physical models of the vocal fold mucosa: Theory and experiment. Lewis P. Fulcher and Ronald C. Scherer (Bowling Green State Univ., Bowling Green, OH 43403)

One of the simpler manifestations of the role of the vertical phase difference in transferring energy from the glottal airflow to the motion of the vocal folds is the surface wave model (SWM) developed by Titze in 1988. He predicted that the phonation threshold pressure (PTP) should decrease as the glottal half width decreases and as the vocal fold thickness increases, and increase as the tissue damping increases, if no vocal tract is present. Further, his treatment of the SWM predicts that for a given glottal half width, converging glottal shapes should have a higher PTP than diverging glottal shapes. Although some of the experiments done with physical models of the vocal fold mucosa support these predictions, the measurements of the angle dependence of the PTP [Chan *et al.*, "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727 (1997)] do not follow the expected trend. A recent re-examination of the SWM found that the diverging-converging question is sensitive to the value of the entrance loss coefficient. Values of the coefficient near 1.37 seem to remove the discrepancy with the data. [Work supported by NIH R01DC03577.]

5aSC29. An acoustic and electroglottographic study of phonation contrast in Yi. Jianjing Kuang (UCLA, Phonet. Lab., Dept. Linguist., Los Angeles, CA 90095-1543)

This study examines the phonation contrast of Yi, a Tibeto-Burman language of Southwestern China with phonation as a phonemic dimension (traditionally described as tense versus lax contrast). The language has seven monophthongs as well as three tones (high, mid, and low). Tense versus lax phonation contrast can apply to all the vowels and two tones (mid and low) in the language. In this study, both acoustic and electroglottographic data were collected with minimal pairs for all combinations of tones and vowels. Extensive acoustic measures were done by VOICESAUCE, including F0, H1, H2, H1-H2, H2-H4, H1-A1, H1-A2, F1, F2, and cepstral peak prominence (CPP); EGG measures were closed quotient (CQ) and peak-velocity (PV). Preliminary data suggest that H1-H2, CPP, CQ, and PV measures can exclusively distinguish the two phonation types. Moreover, EGG data also suggest voice quality interacts with tones, but the patterns vary among speakers. This might due to the different strategies used by speakers to produce the tense versus lax contrast. In addition, formant frequency measures also show salient differences between tense and lax vowels; tense vowels generally have higher F1 than the lax counterparts. All of the facts suggest that phonation contrasts in Yi are realized across different phonetic dimensions.

5aSC30. Acoustic analysis of phonation and tone interactions in Mazatec. Marc Garellek (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543, marcgarellek@ucla.edu)

Jalapa de Diaz Mazatec is unusual in contrasting three phonation types (breathy, modal, and laryngealized) and three tone heights: low, mid, and high, including tone contours. Using the acoustic measures H1-H2, H1-A2, and cepstral peak prominence (CPP), previous research has shown that the three phonation types on mid tones are best distinguished at the beginning of the vowel, with each phonation gradually tending toward breathiness [Blankenship (2002)]. The present study extends previous findings by investigating tone and phonation interactions in over 650 Mazatec words and for a large range of acoustic measures. The speech samples, from the UCLA Phonetics Archive, were produced by six female and eight male speakers and vary across the three phonation types and two tonal groups (non-low and low). Using the voice analysis program VOICESAUCE [Shue *et al.* (2009)], acoustic measures such as harmonic amplitudes, F0, energy, CPP, formants, and bandwidths were extracted. The results generally support previous findings. Several measures also show effects of tone on phonation. In addition, how the phonation types and tone groups differ in timing and the effect of preceding consonants on phonation were investigated. Results show that preceding aspiration changes the timing and degree of phonation. [Work supported by NSF.]

5aSC31. A spectral-slope compensated scale for measuring perception of vocal aperiodicity. Bruce R. Gerratt and Jody Kreiman (Div. of Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794)

The influence of sound intensity on listeners' perception of pitch and of sound frequency on the perception of loudness are well known. Similar interaction effects occur in voice quality perception, but are not well understood. Previous studies suggest that listeners' sensitivity to noise levels in voice depends on the shape of the harmonic spectrum: As the amount of high-frequency harmonic energy present in the voice spectrum increases, listeners require a corresponding increase in noise energy to perceive a constant noise level. These results indicate that perceptually meaningful measurement of spectral noise in voice requires derivation of a spectral-slope-compensated noise scale (similar to equal loudness curves). Based on 12 natural voices, series of stimuli will be created by modifying (1) the noise-to-harmonics ratio (NHR) for different source spectral slopes and (2) the source spectral slope for different NHR values. Just-noticeable differences for each parameter will be plotted as a function of the other, to create a set of equal noise contours reflecting the perceptual interactions of these two acoustic attributes. [Work supported by NIH.]

5aSC32. Voice quality as a cue to location in a speaker's fundamental frequency f0 range: A perceptual study of English and Mandarin. Jason B. Bishop (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095, j.bishop@ucla.edu)

Recent research shows that listeners can locate a specific f0 within a speaker's individual range without prior experience with the speaker's voice [Honorof and Whalen, J. Acoust. Soc. Am. **117**, 2193–2200 (2005)], suggesting that perhaps some acoustic parameters co-vary with f0 range and are available to listeners. Here we tested Honorof and Whalen's hypothesis that this information lies in voice quality. English- and Mandarin-speaking listeners heard brief, steady-state /ɑ/ tokens from various locations throughout the f0 ranges of ten English and ten Mandarin voices and were asked to identify where in the speaker's pitch range a given token came from. To test for listeners' use of voice quality in the task, we compared mixed effect linear models of responses that included three acoustic measures of the stimuli associated with voice quality: H1-H2, H2-H4, and cepstral peak prominence. Preliminary findings show (1) each of the voice quality measures contributed significantly to the models, although they accounted for only a small portion of the variance in responses compared to f0 and (2) models improved significantly when the language of the voice, but not of the listener, was included; this variable represents further language-specific properties of the voices still to be explored.

5aSC33. Online monitoring of auditory feedback is sensitive to language experience. Zhaocong Chen, Peng Liu (Dept. Rehabilitation Medicine, Sun Yat-Sen Univ., Guangzhou 510080, China), Emily Q. Wang (Rush Univ. Medical Ctr., Chicago, IL 60612), Charles R. Larson (Northwestern Univ., Evanston, IL 60208), and Hanjun Liu (Sun Yat-Sen Univ., Guangzhou 510080, China)

It has been demonstrated that vocal responses to pitch perturbations vary as a function of stimulus parameter and can be modulated according to the specific demands of vocal task. The purpose of this cross-language study was to examine whether the online monitoring of auditory feedback is sensitive to language experience during vowel phonation. Native speakers of Cantonese and Mandarin participated in the experiments. They were asked to vocalize a vowel sound /u/ at their conversational pitch, during which their voice pitch feedback was unexpectedly shifted ($\pm$50, $\pm$100, $\pm$200, or $\pm$500 cents, 200 ms duration) and fed back to them over headphones. The results showed that, as compared to previous findings in English speakers [Chen *et al.* (2007)], both Mandarin and Cantonese speakers produced smaller but faster responses to pitch perturbations. In addition, Mandarin speakers produced larger response magnitudes than Cantonese speakers, and the modulation of response magnitudes as a function of stimulus magnitude was observed in Cantonese but not in Mandarin speakers. These findings demonstrate that voice F0 control is language dependent. Further, the dif-

ferent patterns of vocal responses between Mandarin and Cantonese speakers indicate that this highly automatic feedback mechanism works within the specific tonal system of each language.

5aSC34. The role of creaky voice quality in Cantonese tonal perception. Hiu-Wai Lam and Kristine M. Yu (Dept. of Linguist., Univ. of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095, hiuwai208@ucla.edu)

Vance (1976) found a response bias against Tone 4 (mid-low falling) in a tonal perception experiment in Cantonese where synthesized stimuli varied only in F0, and Vance (1977) explained this by suggesting that creaky voice quality is a redundant cue for Tone 4. Indeed, there is evidence that creaky voice quality plays a role in tonal perception: in Mandarin, a language where creak is well-known to be a redundant cue for one of the tones, Tone 3 (low fall-rise), Belotel-Grenié and Grenié (1994) found that creaky instances of Tone 3 were recognized faster than non-creaky instances. The effect of creaky voice quality on the perception of Tones 4 and 6 (mid-low level) in Cantonese will be investigated using a 2AFC identification task of natural stimuli that were elicited in isolation and in connected speech. Variability in creak in the realization of Tone 4 occurred naturally in the elicited stimuli. If creaky voice quality plays a role in tonal perception, we hypothesize that overtly audible creak, as well as low H1-H2, a spectral index of creak, will bias listeners toward identification as Tone 4, as measured by d' scores, and may also speed Tone 4 recognition.

5aSC35. Perceptual similarities between native and non-native tones. Xianghua Wu (Dept. of Linguist., Simon Fraser Univ., 8888 Univ. Dr. Burnaby, BC V5A 1S6, Canada, xianghua_wu@sfu.ca)

This study investigated the effects of L1 and L2 experience on the perceptual assimilation of non-native tones by native Mandarin or Thai listeners. Of these, 32 had 0.5–1.5 years of L2 learning experience while 40 had none. All listeners participated in a tonal assimilation task in which they first identified which tone in Mandarin or Thai sounded most similar to the Thai or Mandarin tone they heard, and then rated its goodness on a five point Likert scale. Stimuli included four Mandarin tones (high level, rising, falling-rising, and falling tones) and five Thai tones (mid, low-, falling, high-, and rising tones) on four monosyllables: /tuo/, /fej/, /kha/, /siau/, and a hum. The Thai listeners, regardless of L2 experience, assimilated more L1 tone categories to L2 than did the Mandarin listeners. Nonetheless, the experienced listeners from the two languages showed a high degree of consistency in terms of which tones were assimilated. The perceived similarities between native and non-native tones were not always predictable from their acoustic similarities, and varied with L1 and L2 experience. Results will be discussed in terms of some well-known perceptual assimilation models, such as PAM. [Work supported by GIS.]

5aSC36. Production and perception of lexical tones in Beijing and Taiwan Mandarin. Chiung-Yun Chang and Robert Allen Fox (SPA Labs, Speech and Hearing Sci. Ohio State, 1070 Carmack Rd., Columbus OH 43210-1002, chang.553@osu.edu)

The acoustic properties of four lexical tones between two regional varieties of Mandarin Chinese, Beijing Mandarin (Putonghua), and Taiwan Mandarin (Guoyu) were examined in terms of duration, pitch contours, and rms amplitude. Tokens included CV and V monosyllables, representing each of the four tones of Mandarin Chinese, and were produced in isolation and in sentence context by 15 adult native speakers. A different durational pattern of citation tones emerged for two dialect variety groups. In Taiwan Mandarin, it was T2>T1>T3>T4; whereas, it was T3>T2>T1>T4 in Beijing Mandarin [Deng *et al.* (2006)]. The durational discrepancy in isolated T3 may be related to different realization of T3 between two dialect groups. While T3 exhibited a falling-rising pitch contour in Beijing Mandarin, it was falling in Taiwan Mandarin. Such dialectal divergence in T3 contours shapes can be verified from the amplitude contours. Furthermore, pitch contours of other tones in two dialects will be compared to see if there is tonetic sound change in other Taiwan Mandarin tones. These surface acoustic variations in linguistically identical categories can result in perceptually ambiguous tones. A gating experiment was utilized to examine how native and non-native listeners adapt to speaker and dialect variability in the stimuli.

5aSC37. Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. Hyunjung Lee and Allard Jongman (Dept. of Linguist., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66044, elvisdj@ku.edu)

The three-way laryngeal distinction among voiceless Korean stops has been well documented for the Seoul dialect. The present study compares the acoustic and aerodynamic properties of this stop series between two dialects, non-tonal Seoul and tonal South Kyungsang Korean. Sixteen male Korean speakers (eight from Seoul and eight from Kyungsang) participated. Measures collected included VOT, f0 at vowel onset and midpoint, H1-H2, and air pressure and airflow. The presence versus absence of tone affects both the acoustic and aerodynamic properties. First, Seoul speakers primarily use f0 to distinguish the three laryngeal gestures of Korean stops, while Kyungsang speakers mainly use VOT. Second, the high versus low tonal contrast for Kyungsang speakers makes f0 an unreliable acoustic cue for the three Korean stops. Third, dialectal differences in VOT to mark the three-way distinction support the notion of a diachronic transition whereby VOT differences between the lenis and aspirated stops in Seoul Korean have been decreasing over the past 50 years. Finally, the aerodynamic results make it possible to postulate the articulatory state of the glottis. Based on the acoustic and aerodynamic results, different phonological representations for the tonal and non-tonal dialects are suggested.

5aSC38. Temporal changes in Mandarin tone due to speaking rate variation. Joan A. Sereno and Hyunjung Lee (Dept. Linguist., Univ. of Kansas, 1541 Lilac Ln., Lawrence, KS 66045, sereno@ku.edu)

The present research investigates the effects of variation in speaking rate on the production of Mandarin tone. Fourteen speakers (6M, 8F) produced 15 syllables, each with the four different Mandarin tones. Speakers produced these syllables at fast, normal, and slow speaking rates in isolation. To induce rate change, the 60 target words were presented on a computer screen at different tempos. Consonant, vowel, and syllable duration were measured as well as turning point (TP) for the contour tones, Tones 2 and 3. TP is the time interval from tone onset to the lowest f0 value. Duration data of each tone revealed significant ambiguity across tones due to speaking rate. Specifically, for the contour tones, results indicated significant overlap in terms of duration and, as speaking rate increased, the TP for Tone 2 decreased to a lesser extent as compared to Tone 3. Moreover, the TP values for Tones 2 and 3 showed a sizable range, with the ambiguous region of overlap observed not at the faster but at the slower speaking rates. These findings will be discussed in terms of the temporal adjustments that occur in production when speaking rate changes. [Research supported by NSF.]

5aSC39. An acoustic and electroglottographic study of Cantonese tone. Kristine M. Yu, Hiu-Wai Lam, and Shing-Yin Li (Dept. of Linguist., Univ. of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095, krisyu@humnet.ucla.edu)

Cantonese is a tone language with six tones (traditionally described as high level, high rising, mid level, mid-low falling, mid-low rising, and mid-low level), and also three "stopped" tones in CVC words, not studied here. The interest of this study was (1) how different acoustic cues can be used to classify the different tones, particularly in connected speech and (2) the interaction of voice quality and tone in Cantonese. Twelve speakers were recorded producing minimal pairs as in Wong [(2006)] in isolation, in isolated disyllables, and in sentence-medial disyllables, where the disyllables ranged over all possible 36 bitones in the language; all speakers also made electroglottographic (EGG) recordings. Acoustic measures were F0, F0′, cepstral peak prominence, and harmonic amplitudes $H1^*$ and $H2^*$, $H1^*$-$H2^*$, $H1^*$-$A1^*$, $H1^*$-$A2^*$, $H1^*$-$A3^*$, and $H2^*$-$H4^*$. EGG measures were closed quotient and peak of increasing contact. Measures were made automatically using VOICESAUCE and EGGWORKS. Preliminary results indicate variability in

the presence of creak in the realization of the lowest tone (mid-low falling) as well as variability in the presence of creak at the low f0 regions of the rises.

5aSC40. Identification of prominence and discrimination of pitch patterns by Japanese and American listeners. Irina A. Shport and Susan Guion-Anderson (Dept. of Linguist., Univ. of Oregon, 1290 Eugene, OR 97403, ishport@uoregon.edu)

This study examines how native language shapes the perception of prominence in three-syllable nonce words *nenema* with F0 patterns varying in F0 peak alignment and F0 fall. F0 is the fundamental cue to perception of pitch accent in Japanese, in which the accent location and accent type (accented versus unaccented phrases) are mainly defined by the relationship between the F0 peak, which is associated with the accented syllable, and the F0 fall, which follows the peak. In English, F0 fall is not considered to be a cue to stress. In two experiments, the alignment of the F0 peak (eight locations) and the F0 fall (no fall, moderate fall, steep fall) were manipulated. In the identification task, participants were asked to decide whether the first or the second syllable sounded more prominent to them. In the discrimination task, participants decided whether the pitch patterns of two words were the same or different. Japanese listeners were expected to be more sensitive to the F0 fall than American listeners for all F0 peak locations. Difference in the perception of peak location between the groups was also predicted, as a reflection of peak delay typical to the native language.

FRIDAY MORNING, 23 APRIL 2010 ESSEX A/B/C, 7:35 A.M. TO 12:00 NOON

Session 5aSP

Signal Processing in Acoustics, Underwater Acoustics, and Animal Bioacoustics: Classification Methods in Acoustics and Non-Gaussian Noise I

R. Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Brett Bissinger, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Chair's Introduction—7:35

Invited Papers

7:40

5aSP1. Statistical pattern classification: A review and some current problems/paradigms. David J. Miller (Dept. of Elec. Eng., The Penn State Univ., Rm. 227-C EE West Bldg., Univ. Park, PA 16802)

Unsupervised clustering is the ability to automatically partition a set of data patterns into meaningful groups without prior knowledge of the groups or their number. Supervised classification is the ability to automatically recognize a data pattern as an instance from one of a known set of classes. These problems are fundamental to a variety of application domains, including scientific (e.g., bioinformatics), engineering (e.g., speech recognition), business (e.g., marketing and document clustering), and military (target detection). In this talk, we first review the basic supervised and unsupervised classification problems, standard solution methodologies, and how to characterize their performance. We then discuss some recent problem variants and associated paradigms, including semisupervised learning, unsupervised clustering in high-dimensional feature spaces, and decision fusion techniques.

8:00

5aSP2. Target/clutter discrimination using Bayesian active/passive data fusion. Brian R. La Cour, Jason M. Aughenbaugh, Bryan A. Yocom, and Thomas W. Yudichak (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Discriminating targets of interest from background clutter is a key challenge in undersea active sonar. This is particularly true in littoral areas, where scattering from the bottom, motion in the water column, and activity on the surface all contribute to produce echoes which are difficult to distinguish from targets of interest. Statistically, these effects manifest themselves in non-Rayleigh, heavy-tailed amplitude distributions. An approach to target/clutter discrimination is described which uses the complementary information from both active and passive acoustic sensors to facilitate this task. The method uses an efficient, grid-based Bayesian track-before-detect scheme to combine data from the two types of sensors by carefully modeling the effects of array processing, replica correlation, normalization, and clutter statistics. Representing measurements and uncertainty in terms of likelihood functions then provides a common framework for fusion. In this manner, active returns with coincident and appropriate passive signals are given more credence, while the presence of only one or the other, while perhaps suggestive, is not as compelling. The presentation will give a general overview of the approach, and an example will be used to illustrate the potential power of using passive data to mitigate active clutter. [Work supported by the Office of Naval Research.]

8:20

5aSP3. Moments in history. Leon Cohen (Dept. of Phys., Hunter College of CUNY, 695 Park Ave., New York, NY 10065)

We discuss how moments and functions of moments have been historically used to characterize and classify distributions, as well as to construct distributions. Among the methods that have been devised are the Gram-Charlier and Edgeworth series, which are methods used to improve on the Gaussian distribution when the moments deviate from Gaussian. We show how these series can be generalized and how any two distributions can be transformed into each other by systematically changing the moments. In this manner, one can develop families of distributions. We illustrate the methods by applying them to pulse propagation in a dispersive medium where the moments can be deterministic or stochastic. [Work supported by the Office of Naval Research.]

8:40

5aSP4. Computationally enabled alternatives to Gaussian classification and tracking—A survey. John R. Sacha (Appl. Res. Lab., Penn State, P.O. Box 30, State College, PA 16804, jrs9@psu.edu)

The normal distribution is a good model of physical phenomena in numerous situations where central limit theorem effects come into play. It has a prominent place historically because the many special properties of the Gaussian PDF often yield mathematically elegant closed-form solutions to problems. This role was especially important when numeric calculation was expensive. Normality is not a valid assumption in many applications of modern acoustical signal processing, perhaps especially involving classification and tracking. For pattern recognition, Gaussian classifiers are a straightforward methodology requiring mean and covariance estimates of the class distributions. However, in many cases, feature PDFs themselves are not multivariate-Gaussian; even if they are, estimating parameters in large feature spaces with limited data is difficult. In tracking, the Kalman filter is the canonical formulation, but it requires normally distributed process noise and measurement errors. The advent of cheap computation power has made practical a host of alternative models. In classification, this includes empirical PDF estimation via binning, fuzzy logic, neural nets, support vector machines, featureless classification, Bayesian belief nets, and decision trees. For tracking, particle filters are used to model densities as clouds of point estimates; although computationally expensive, they admit arbitrary densities and are amenable to parallel implementation. [Portions of this work were sponsored by ONR.]

9:00

5aSP5. Estimation of densities and information measures: An overview and some recent results. Sanjeev R. Kulkarni (Dept. of Elec. Eng., Princeton Univ., Princeton, NJ 08544)

A fundamental problem that arises in many applications is to estimate an unknown probability distribution from a set of observations drawn according to this distribution. Of particular interest is the case of non-parametric estimation in which little or nothing is known about the underlying distribution. Surprisingly, consistent estimators can be obtained even in this case and the resulting methods are called universal. A related problem is to estimate various information measures based on observations drawn one or two unknown distributions. This talk provides an overview of universal density estimation and the problem of estimating information measures. Some recent results and applications are also described.

9:20

5aSP6. Classification of non-Gaussian reverberation as a mixture. Nicholas P. Chotiros (Appl. Res. Labs., The Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758, chotiros@arlut.utexas.edu)

There are a few different ways to obtain non-Gaussian statistics, one of which is through the heterogeneity of the seafloor. It is often the case that the seafloor is a patchwork of different bottom types. Measurements of seafloor reverberation, using omni-directional sources and receivers, often show Gaussian statistics, because it is the result of the superposition of a large number of random contributions, satisfying the central limit theorem. For sonar systems that have the spatial resolution to resolve the patches, the statistics of the reverberation will change from one resolution cell to the next. Due to positional inaccuracies associated with most sonar systems, it is often not possible or feasible to separate the reverberation from each bottom type. When taken as a whole, the reverberation will have non-Gaussian statistics. The statistical properties will be explored. [Work supported by the Office of Naval Research, Ocean Acoustics Program.]

9:40

5aSP7. Characterization of non-Gaussian, bistatic, acoustic clutter on the Malta Plateau. John R. Preston (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

Sonar clutter is one of the primary limitations to active ASW. This work focuses on statistical analysis of clutter from some bistatic measurements. Non-Gaussian characterizations of bistatic clutter from a mud volcano field, a shipwreck and a high clutter patch of the Malta Plateau are presented. The received data are taken from the five octave research array (FORA) that has been used to collect extensive monostatic and bistatic data in two recent sea trials on the Malta Plateau off Sicily called Clutter 07 and BASE 07. This work uses data from the above mentioned sea trials to characterize non-Gaussian behavior of bistatic clutter from pulsed sources in the 8009576;3500-Hz band and to compare bistatic clutter with monostatic clutter. K-distributions with their shape and scale parameters are used to describe non-Gaussian behavior together with the models of Abraham and Lyons to infer scatterer densities. The ability to geo-reference key statistical measures of clutter can allow CFAR processors to adaptively set thresholds and reduce false alarms. Examples are shown to demonstrate this. Also included are presentations of the shape parameter vs bistatic aspect angle where appropriate. [Work supported by ONR Code 321US].

10:00—10:20 Break

10:20

5aSP8. Active sonar clutter classification using higher order moments. James M. Gelb and Andrew W. Oldag (Appl. Res. Labs., Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758, gelb@arlut.utexs.edu)

The statistics of normalized matched-filter echoes from an active sonar system operating in a myriad of oceanic environments has been studied extensively for three broad clutter classes including using low-order cumulants to classify subregions of the data [Gelb *et al.*, Proceedings of the ISURC (2008) and references therein]. That work compared empirical distributions to parametric models (e.g., the *K* distribution and the generalized Pareto distribution). A report on extensions of this work is presented including studies of the accuracy of analytic parameter estimation methods and the efficacy of using higher order moments in the classification process. For each class, with increasingly heavy non-Rayleigh distributed tails, comparisons are made of brute force parameter estimation with the use of analytic estimators. Additionally, comparisons of higher order moments (including skew and kurtosis) computed from the data are made with analytic fits to the data. Using a feature-based classifier, the gains of using increasingly higher order moments are assessed. [Work sponsored by the Office of Naval Research (ONR).]

10:40

5aSP9. Improving anti-submarine warfare tracking performance by incorporating classification information. William H. Mortensen, David W. Krout, and Jack McLaughlin (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105-6698, william5@u.washington.edu)

Prior research has shown that more accurate classification of contact amplitudes can improve tracking performance. In this work, we augment standard, contact-based trackers with a classifier run on features of the received time series from which the contacts were extracted. Ground truth information from benchmark datasets is used in a flexible simulation framework built around the sonar simulation toolset (SST) to generate simulated target and clutter returns. The simulated returns are used as input to the classifier and the contacts from the benchmark datasets as input to the tracker. The classification information provides an additional input to the association step in the probabilistic data association (PDA) and joint PDA trackers, and to the probability of target for each contact in the Bayesian tracker. The results show that even relatively poor classification can make a noticeable improvement in tracking performance. [This work was funded by the U.S. Office of Naval Research, Contract No. N00014-01-G-0460, Delivery Order #36.]

11:00

5aSP10. Model-based detection of buried objects. Edmund J. Sullivan (EJS Consultants, 46 Lawton Brook Ln., Portsmouth, RI 02871, paddypriest@aol.com) and Ning Xiang (Rensselaer Polytechnic Inst., Troy, NY 12180)

An approach to the detection of buried objects is to excite the ground with low-frequency, high-energy sound waves, which then excites a resonance in the buried object. The ensuing vibration causes a detectable signal on the surface of the ground, which can be detected using a laser doppler vibrometer (LDV). The original detection technique used a sliding bandpass filter to process the scattered LDV energy, providing an energy map of the area scanned by the LDV, which indicates the location of the object. The performance of detection is often limited by speckle noise, a type of noise arising from the coherent nature of the laser beam. A more recent technique utilizes an autoregressive model of this noise. This leads to an inverse filter that whitens the noise. Upon the appearance of any target data in the signal, a whiteness test indicates a detection. This approach has demonstrated improvement over the bandpass filter approach. This paper demonstrates a further improvement by augmenting the prewhitener with a model of the mine itself. This provides significant improvement by both enhancing the mine signal and improving the detection performance. Experimental results are shown.

11:20

5aSP11. Sonar waveform design for detection of elastic objects. Brandon Hamschin (Dept. of ECE, Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261, bmh52@pitt.edu) and Patrick Loughlin (Univ. of Pittsburgh, Pittsburgh, PA 15261)

Animals that navigate and hunt by echolocation, such as some bats and marine mammals, have been observed to change their sonar pulse depending on the environment, as well as during hunting. It has become of interest to incorporate these strategies into man-made sonar waveform and receiver design. We examine the benefits of optimal waveform design versus transmitting a linear FM waveform for detecting elastic objects. Performance loss suffered by assuming a point target is also examined. Our approach utilizes a method recently proposed by Kay to design the optimal power spectrum of the transmit waveform. Because there is an unlimited number of waveforms with the same power spectrum, we further impose a time domain constraint, in terms of the signal duration, to obtain a unique optimal waveform. [Work supported by ONR 321US.]

11:40

5aSP12. Error bounds for classifying targets in non-Rayleigh clutter. Douglas A. Abraham (CausaSci LLC, P.O. Box 5892, Arlington, VA 22205, abrahad@ieee.org)

False alarms in active sonar systems are often represented statistically as having a probability density function (PDF) with tails heavier than the traditionally assumed Rayleigh PDF. Distributions such as the Weibull, *K*, and Poisson–Rayleigh have been used to represent such non-Rayleigh clutter and to derive the associated probabilities of false alarm and detection, the latter for standard target models (e.g., nonfluctuating and fluctuating). In this presentation, the probability of at least one misclassification per ping (P_c) is evaluated when classifying the standard target types against non-Rayleigh clutter having equal average power. The Bhattacharyya bound is used as an upper bound for the Bayesian error probability (P_e) of a classifier operating on a single cluster and combined with the average number of clusters per ping (N_c) to form P_c. As expected, P_c increases as the clutter becomes more non-Rayleigh. Clutter statistics are seen to affect P_c through both P_e and N_c. However, the effect of clutter statistics on N_c is significantly greater than on P_e. Increasing bandwidth reduces P_c (and therefore P_e) when its impact is modeled as an increase in the number of independent samples available for classification. [Work sponsored by the Office of Naval Research under Contract No. N00014-09-C-0318.]

Session 5pAA

Architectural Acoustics: Potpourri, Measurements, Equations, and Case Studies

Kenneth W. Good, Chair
Armstrong World Industries Inc., 2500 Columbia Ave., Lancaster, PA 17603

Contributed Papers

1:00

5pAA1. Comparison of two just-noticeable-difference test methods for clarity index (C80). Clothilde B. Giacomoni, Christopher M. Jasinski, Robert D. Celmer, and Michelle C. Vigeant (Dept. Mech. Eng., Acoust. Prog. and Lab., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

The just noticeable difference (JND) or smallest detectable increment of clarity index (C80) has been investigated due to the lack of consensus in the existing literature. The purpose of this study was to determine how the JND of C80 varies as a function of the test procedure. Test signals, with varying amounts of clarity, were generated and combined with short anechoic recordings of orchestral music. The testing took place in the University of Hartford's anechoic chamber, and the signals were played back over eight spatially arranged loudspeakers. Two testing methods were compared, which both consisted of the subject hearing two signals, A and B, and deciding if the signals were the same or different in terms of clarity. For Test Method 1, the subjects were required to listen to all of signal A and then all of signal B before selecting their response. For Test Method 2, the subjects were allowed to switch between signals A and B in real-time. The difference in the JND of C80 for the two test methods, along with a comparison to previously published results, will be discussed. [Work is supported by the Paul S. Veneklasen Research Foundation.]

1:15

5pAA2. The effect of motif length in reverberation-time listening tests using the double-blind testing method ABX. Scott S. Edwards, Daniel A. Ignatiuk, Robert D. Celmer, and Michelle C. Vigeant (Dept. Mech. Eng., Acoust. Prog. and Lab., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

Subjective listening tests are often conducted to determine the relationship between an objective parameter and human perception. These tests can consist of auralizing measured or predicted room impulse responses (RIRs) that have been convolved with anechoic recordings. However, no previous research exists that provide guidelines in terms of how auditory memory affects the chosen length of the musical motif convolved with the RIR. Research in the field of auditory memory suggests that there is both short auditory memory of approximately 200 ms and long auditory memory lasting several seconds [Cowan, Psychol. Bull. **96**, 341–370 (1984)]. A study was carried out using two motifs with three different lengths: 5, 7, and 10 s, where the successively longer motifs contained the same passages as the shorter ones. A relatively straightforward parameter of reverberation time was the independent variable, and signals were created with reverberation times between 1.0–1.5 s. The subjects were presented three signals, A, B, and X, where the difference in reverberation time ranged between 0.3–0.5 s. The subjects were required to identify which signal A or B matched the test signal X. The effect of motif length on the percent correct will be discussed. [Work is supported by the Paul S. Veneklasen Research Foundation.]

1:30

5pAA3. Modeling specular reflection in enclosures with energy-intensity boundary elements and an iterative relaxation method. Krista A. Michalis, Donald B. Bliss, and Linda P. Franzoni (Pratt School of Eng., Duke Univ., 1111 Hudson Hall, Durham, NC 27708, kam49@duke.edu)

Steady-state sound fields in enclosures with specular reflection boundaries are modeled by an energy-intensity boundary element method using uncorrelated broadband directional sources. The specular reflection case is solved using an iterative relaxation method extended from a three dimensional (3-D) diffuse reflection solution. When chosen properly, few directivity harmonics are required. Starting with a diffuse calculation, higher harmonics are estimated by post-processing, and the original diffuse influence matrix is refined accordingly and converges in few relaxation steps. The method is higher-order accurate and similar to the diffuse case in that it includes strict enforcement of energy conservation. The method is computationally efficient, making it useful for the design of acoustic spaces. This new 3-D energy-intensity boundary element method utilizes an absorption scaling theory, allowing for direct solution of the element power distribution that describes the primary spatial variation caused by the effects of specular reflection, the absorption dependence on incidence angle, and the nonuniform spatial distribution of absorption and input power in the enclosure. Sample results show that differences between diffuse and specular solutions are often fairly small. The iterative relaxation method, with a limited number of specular harmonics, appears to provide a reasonable approach in practice. [Sponsored by NSF.]

1:45

5pAA4. Control algorithms for adaptive time-frequency test signal synthesis for architectural acoustics measurements during music concerts. Gang Ren (Dept. of Elec. and Comput. Eng., Edmund A. Hajim School of Eng. and Appl. Sci., Univ. of Rochester, Rochester, NY 14627), Justin Lundberg (Univ. of Rochester, Rochester, NY), Dave Headlam, and Mark F. Bocko (Edmund A. Hajim School of Eng. and Appl. Sci., and Appl. Sci., and Univ. of Rochester, Rochester, NY)

A system was proposed by G. Ren *et al.* [158th Meeting of the ASA (2009)] to perform field acoustics measurements during ongoing music concerts. It addresses the long-standing problem of performing architectural acoustic measurements with an audience present. In this system, an adaptive time-frequency synthesis algorithm was employed to generate test signals with time-frequency energy distributions that conform to predicted "masked measurement vacancies" to ensure that the acoustic response of the test signals can be separated from the music and that the test signals are inaudible to the audience by taking advantage of auditory masking phenomena. In the present paper, approaches utilizing higher-level structure of music are exploited to reduce the number of conflicts between the test signals and the music. Three approaches are described: a music theory based probabilistic prediction algorithm, using prior knowledge of the score, and the third exploits the availability of previous recordings of the music being performed. These control algorithms are adapted to architectural acoustic measurement procedures conforming to ISO3382-1/2(2009). The proposed system achieves similar performance compare to conventional acoustic test procedures, which emit loud test signals in empty music halls.

2:00

5pAA5. Improving FSTC (field sound transmission class) and FIIC (field-impact isolation class) of standard wall and floor/ceiling configurations. Bonnie Schnitta, Melissa Russo, Trish Kern, and Greg Greenwald (SoundSense, LLC, 46 Newtown Ln., Ste. One, East Hampton, NY 11937)

Every acoustic floor or wall design requires an analysis in order to provide a sustainable design. This is true whether the design is a single family residence with the goal of achieving a desired level of quietude, or a multi-family, or commercial dwelling required to meet various codes and/or standards. Since IIC and STC measurements taken in the field produce a lower STC or IIC than laboratory controlled values, this presentation will begin by reviewing the variables that are not present in the controlled laboratory setting for their acoustic impact. This presentation will further provide field readings of well known documented basic floor and wall configurations. With this foundation standard floor and wall configurations will be reviewed for their FSTC and FIIC improvement with standard readily available acoustic materials to address variables consistent with the degradation of the configuration's acoustic efficacy. The FIIC and FSTC values of the same floors and walls will be further compared to walls and floors with unique and innovative installation methodologies and configuration upgrades. The summary will be a detailed approach to the best field practices and innovative configuration upgrades, inclusive of a few new patented and patent pending products, for optimum field acoustic efficacy.

2:15

5pAA6. Modeling diffuse and specular reflection surfaces and their effect on broadband interior sound field level and intensity flow. Donald B. Bliss, Linda P. Franzoni, and Krista A. Michalis (Mech. Eng., Duke Univ., Durham, NC 27708, dbb@duke.edu)

For steady-state broadband sound fields in enclosures, surfaces may be modeled as diffuse reflectors, specular reflectors, or some combination. Modeling of reflection surfaces when using energy-intensity based boundary elements is reviewed, and the proper energy-intensity source characteristics are derived. These characteristics depend on local correlation effects, the angular distribution of the incident field, and the angular dependence of reflection coefficient. Specular reflection surfaces subjected to fairly random reverberant sound fields exhibit behavior similar to diffuse reflection surfaces to lowest order. Simple solutions are obtained to illuminate the basic difference in interior sound fields caused by different reflection types. For example, random incidence sound propagating down a long corridor with non-absorbing walls exhibits different behavior depending on whether the surfaces are diffuse or specular reflectors. The diffuse case exhibits a spatial gradient in mean-square pressure along the corridor, whereas the specular case gives a uniform field. These solutions are also used to clarify the relationship between the flow of intensity and the gradient of mean-square pressure. This relationship has been represented simplistically and incorrectly in some energy based methods. The importance of benchmarking energy based methods against analytical solutions, rather than experimental data, is stressed.

2:30

5pAA7. Architectural impacts on speech intelligibility in atrium spaces: Two case studies. Julie E. Fischer (Shen Milsom & Wilke, LLC, 3300 N Fairfax Dr., Ste. 302, Arlington, VA 22201, jfischer@smwllc.com)

Case study reviewing the speech transmission index and voice intelligibility design and implementation of two atrium spaces where dedicated public address systems were implemented. The architectural design and selection of finishes has a great impact on the capability of any public address loudspeaker system distinct from any system performance characteristics. The presentation will review the design process, acoustical modeling, and testing results outlining the issues associated with meeting the NFPA 72 Common Intelligibility Score suggested goals. Architectural design requirements such as day-lighting and views are increasing large areas of glass in atrium spaces. These designs are making intelligibility increasingly difficult and the electronic systems and loudspeaker designs can only accomplish so much without architectural acoustical finishes as part of the design.

2:45

5pAA8. Characterization of classroom noise and noise control techniques and their effect on speech comprehension while learning. Sooch SanSouci, Line Guerra (Dept. of Res., AiA-Audition, Intelligibility, Acoust., 47 rue Le Corbusier, 92100 Boulogne, France, s.sansouci@aiacoustique.com), and Dick Campbell (Bang-Campbell Assoc., Falmouth, MA)

Classrooms have been shown to be prone to elevated occupied noise levels that reduce speech comprehension and inhibit learning. This presentation discusses a current study that includes three parts: (1) the characterization of masking invoked by noise commonly found in today's classrooms, (2) the formulation of a metric that intends to associate masking risk to noise features, and (3) the attempt to quantify any increase in speech comprehension resulting from interior acoustic designs that aspire to reduce noise in occupied classrooms. Calibrated recordings made during classroom activities in 9 schools are post processed to allow noise characteristics to be rated based on attributes related to the noise source or type, duration, rate of recurrence, spectra, level, envelope, and peak energy. Binaural and monaural recordings are compared. Speech to noise ratios are statistically weighted over various time frames and activities in an attempt to refine reproducibility among different activities. All recordings were made in occupied rooms both before and after the various treatments were integrated. The results suggest that there are misconceptions in the literature in terms of today's classroom design trends and a few novel principles emerged as being highly effective.

Session 5pAB

Animal Bioacoustics and Psychological and Physiological Acoustics: Auditory Attention, Learning and Memory: From Neurons to Behavior

Cynthia F. Moss, Chair
Univ. of Maryland, Dept. of Psychology, College Park, MD 20742

Invited Papers

1:00

5pAB1. Focusing, maintaining, and switching attention. Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., 677 Beacon St., Boston, MA 02215)

Humans and other animals have an amazing ability to selectively attend to whatever sound source is most relevant, a task that requires the brain to separate a sound mixture into distinct perceptual objects. Results from a number of recent behavioral and neuroimaging studies from our laboratory and others have begun to tease apart the processes and mechanisms that enable us to select an important source from a sound mixture and attend to it as its meaning unfolds, as well as to switch attention if the need arises. As reviewed in this talk, the processes of focusing, maintaining, and switching auditory attention involve dynamics that directly impact our ability to understand sounds in complex settings. For instance, there is a cost associated with the process of disengaging attention from one source and focusing on another. On top of this switching cost, there is a benefit of maintaining attention to an ongoing source that yields improvements in performance over time. These processes interact with the way we store and remember signals and directly influence how we function in complex auditory scenes, especially social settings like the Speech Communication Poster Session or the Thursday Evening Buffet Social.

1:25

5pAB2. The role of innate and attentional mechanisms in parsing complex acoustic scenes: A neural and behavioral study. Mounya Elhilali (Dept. of Elec. & Comput. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218, mounya@jhu.edu)

The mechanisms by which a complex auditory scene is parsed into coherent objects depend on poorly understood interactions between task-driven and stimulus-driven attentional processes. We use a simultaneous psychophysical-neurophysiological experimental paradigm to manipulate human listeners attention to different features of auditory scenes. In a series of experiments, our findings reveal a role of attention in enhancing the sustained neural representation of the foreground. This enhancement, in both power and phase coherence, originates in auditory cortex, occurs exclusively at the frequency of the target rhythm, and is only revealed when contrasting two attentional states that direct subjects focus to different features of the acoustic scene. It also interacts with innate processes of the auditory system, particularly its differential sensitivity to temporal dynamics of sounds. These results have substantial implications for models of foreground/background organization and mechanisms mediating auditory object formation.

1:50

5pAB3. Intention and attention: Top-down influences on the representation of task-relevant sounds. Jonathan B. Fritz (Inst. for Systems Res., ECE, Univ. of Maryland, College Park, MD 20747, ripple@isr.umd.edu), Stephen V. David, Daniel Winkowski, Pingbo Yin (Univ. of Maryland, College Park, MD 20747), Mounya Elhilali (Johns Hopkins Univ., Baltimore, MD 21218), and Shihab A. Shamma (Univ. of Maryland, College Park, MD 20747)

To explore the role of top-down projections from frontal cortex (FC) to auditory cortex (AC), which may play a role in adaptive reshaping of A1 receptive fields to attended acoustic stimuli during behavior, simultaneous recordings were made from single neurons in FC and AC in behaving ferrets, trained on multiple auditory detection and discrimination tasks in positive and negative reinforcement paradigms. Performance required selective attention to different salient spectral frequency and/or temporal cues. Previous observations revealed task-specific transformations in receptive fields in AC [Fritz *et al.* (2003); (2005); (2007); (2009); Atiani *et al.* (2009)]. In contrast, FC neurons showed recognition responses, which categorically distinguished between acoustic foreground and background stimuli. FC responses to targets were often independent of the acoustic properties of the target and thus encoded an abstract representation of the class of target stimuli [Fritz *et al.* (2009)]. Stimulation of FC, paired with tones, lead to receptive field transformations similar to those observed in behavior. These results emphasize the importance of interactions between multiple areas during selective attention, and the tight coupling of target recognition and auditory attention that enhances auditory cortical filters for attended acoustic stimuli, thus creating a functional representation of task-salient sounds during behavior. [This work was supported by grants from NIDCD, NIH.]

2:15

5pAB4. Learning alters the coding capacity of single neurons and populations in the auditory forebrain. James Jeanne (Neurosci. Graduate Program, Univ. of California San Diego, La Jolla, CA 92093), Tatyana Sharpee (Salk Inst. for Biological Studies, La Jolla, CA 92037), and Timothy Gentner (Univ. of California, San Diego, La Jolla, CA 92093)

To control adaptive behaviors, vertebrate neural systems must extract information from large numbers of diverse, physically complex signals in the natural world. Sensory experience helps neural systems to meet this challenge by evoking long-lasting changes that bias cortical circuits toward representations of behaviorally relevant signals. Such circuits may encode stimuli using highly selective individual neurons or by larger populations of neurons, thereby greatly expanding coding capacity. I will describe studies investigating the representation of a complex acoustic communication signal, birdsong, across multiple forebrain regions that are analogous to secondary auditory cortices in mammals. I will show that single neurons and populations of neurons in these regions are tuned to the complex spectro-temporal features in songs that birds have learned to recognize. These results demonstrate that learning mechanisms can act and different organizational levels in the brain, and provide biological support for specific population coding schemes.

2:40

5pAB5. Top-down vocal feedback control in active hearing. Xiaoqin Wang (Dept. of Biomedical Eng., Johns Hopkins Univ., 720 Rutland Ave., Baltimore, MD 21205, xiaoqin.wang@jhu.edu)

Speaking is a sensory-motor process that involves constant self-monitoring to ensure accurate vocal production. This monitoring of vocal feedback allows a speaker to quickly adjust speech production to correct perceived errors between intended and actually produced vocal sounds. The self-monitoring in speaking is crucial for learning to speak native as well as foreign languages. An important behavior in vocal feedback control for both human speech and animal vocalizations is the compensatory change in speech or vocalizations (e.g., pitch, frequency, and intensity) when there is a mismatch between intended and perceived vocal sounds. Such a behavior requires mechanisms for continuously monitoring auditory feedback during vocal production. We have found that disruption of auditory feedback during vocalization alters coding properties of auditory cortex neurons in marmosets (a highly vocal primate species). Furthermore, when marmosets compensate for changes in vocal feedback, there are corresponding changes in their cortical neural activity. These findings suggest that the neural network underlying self-monitoring of vocal production likely consists of both sensory processing and top-down modulations via higher-cortical areas involving planning and memory.

3:05—3:20 Break

Contributed Papers

3:20

5pAB6. Production and perception of sequential patterns in budgerigar (*Melopsittacus undulatus*) warble: Evidence for a magic number. Hsiao-Wei Tu and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, hsiaowei@umd.edu)

The warble song of budgerigars (*Melopsittacus undulatus*) is composed of a large number of elements uttered in streams lasting from a few seconds to a few minutes without obvious repetition of particular patterns. Previous work showed that warble elements can be classified into eight acoustic categories for which budgerigars have corresponding perceptual categories. Here, we analyzed long sequences of natural warble to determine the proportion of these categories in warble as well as any sequential patterns of these categories in warble. Results showed that elements were not randomly arranged and that warble has at least a fifth-order Markovian structure. This suggests that budgerigars may have a motor control of approximately five elements during the production of warble. Investigations of the budgerigars' ability to perceive warble sequences of various lengths showed convergent evidence that budgerigars are able to master a novel sequence between four and seven elements in length. In other words, budgerigars may have an attention span of about five elements which may be a limit of their capacity of neural processing. Through gradual training with chunking (about five elements), birds are able to learn sequences up to 50 elements. [Work supported by NIH/NIDCD R01DC000198 to R.J.D.]

3:35

5pAB7. Perception of alterations in natural budgerigar (*Melopsittacus undulatus*) warble: Implications of animal "syntactical capability." Hsiao-Wei Tu and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, hsiaowei@umd.edu)

The warble song of male budgerigars (*Melopsittacus undulatus*) is an extraordinarily complex, multisyllabic, learned vocalization that is produced continuously often for minutes at a time. Previous work has shown that warble elements form acoustic and perceptual categories for budgerigars and these birds can detect ordering changes in artificial sequences of warble elements. Using operant conditioning and a psychophysical paradigm, we examined the sensitivity of budgerigars for detecting different types of insertions (e.g., targets) in a running background of warble up 1000 elements in length. When the inserted targets are warble calls taken directly from the background, budgerigars show a species-specific ability to detect them solely based on sequence violations in the natural ordering of warble elements. Moreover, budgerigars, but not other species, are especially sensitive to temporally reversed warble elements inserted in natural warble sequences, indicating that the acoustic details of warble elements are also perceptually significant to budgerigars besides sequential cues. Although it is still unclear whether budgerigars perceive their warble as a rule-govern sequence or a pattern-based vocalization, the findings here open the door to studies of serial order learning in a natural, non-human communication system. [Work supported by NIH/NIDCD R01DC000198 to R.J.D.]

3:50

5pAB8. Bioacoustic and behavioral correlates of spatial memory in echolocating bats. Jonathan R. Barchi (Dept. of Neurosci., Brown Univ., 185 Meeting St., Providence, RI 02912, barchi@brown.edu), Jason E. Gaudette, Jeffrey M. Knowles, and James A. Simmons (Brown Univ., Providence, RI 02912)

Echolocating bats face a unique orientation problem due to their reliance on echolocation over visual orientation. Each observation (sonar vocalization) only contains information about objects in front of and within at most 5–10 m of the animal. Because their behavior extends over much larger distances, spatial memory might be particularly important during foraging activity and free flight to integrate the shorter views provided by sonar. To investigate memory for object locations, *Eptesicus fuscus* were allowed to fly freely in an instrumented flight room populated with sparse obstacles (hanging chains). Bat position and vocalization were monitored with a stereo-registered pair of thermal cameras and an array of ultrasonic microphones synchronized to the video. Data were analyzed for correlates of spatial memory-flight path dynamics, head aim, and temporal structure of echolocation signals. Data recorded from bats free-flying in the chain array are consistent with memory of obstacle locations and high-level planning of a flight path through the entire space. These data reflect the behavior of bats flying in natural conditions. [Work supported by ONR and NSF.]

4:05

5pAB9. Effect of ototoxic drugs on Rana catesbeiana tadpole orientation behavior and caspase-3 expression. Erika Alexander, Brian Schmidt, and Andrea Simmons (Dept. of Psych., Brown Univ., 89 Waterman St., Providence, RI 02912, erikaealexander@gmail.com)

Treatment with the ototoxic antibiotic gentamicin results in specific damage to hair cells of the inner ear and possibly of the lateral line system. In amphibians, damaged hair cells may spontaneously recover. Rana catesbeiana tadpoles were treated with different dosages of gentamicin solution, including a nontreated control, for 24 h. Animals were tested behaviorally and then sacrificed either immediately or after a 7 day recovery period. Their brains were processed for immunohistofluorescent labeling of caspase-3, a marker for apoptosis. Behavioral changes associated with gentamicin treatment included a disruption of balance, altered swimming, and an inability to detect water currents. Animals treated with gentamicin showed increased caspase-3 expression in multiple brain areas compared to non-treated controls. Animals allowed a 7 day recovery period that showed lower caspase-3 label across brain areas than animals sacrificed immediately after treatment. These results suggest that ototoxic damage is associated with increased caspase-3 expression in the central nervous system. In tadpoles, affected behaviors undergo spontaneous recovery, which is itself correlated with decreases in caspase-3 expression. [Work supported by NSF GRFP, NIH, and RI Space Grants.]

FRIDAY AFTERNOON, 23 APRIL 2010 LAUREL C/D, 1:00 TO 5:00 P.M.

Session 5pBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasonic Characterization of Bone II

Keith A. Wear, Chair
Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993-0002

Invited Papers

1:00

5pBB1. Ultrasonic characterization of cancellous bone. Biot approach: Pros and cons. Pascal Laugier (Laboratoire d'Imagerie Parametrique, Univ. Pierre et Marie Curie-CNRS,)

Reliable quantification of microstructural and mechanical bone properties remains an open issue with relevance for the diagnosis of bone quality disorders, such as osteoporosis. The reconstruction of such parameters from nondestructive testing based on ultrasound testing and model-based solution of the identification inverse problem have been suggested by several investigators as novel techniques with high potential not only due to the reduced cost and its non-ionizing nature, but for the direct relationship and sensitivity of the propagation characteristics estimates to relevant bone properties that determine mechanical strength, Recently, approaches taking advantage of the poroelastic nature of cancellous bone have been developed. These approaches assume that the Biot theory is a valid model of wave propagation and are aiming at either (i) recovering the properties of the fast and slow waves or (ii) directly reconstructing the bone properties that govern Biot's theory. However, the most recent literature reveals inconsistencies between Biot model and experimental results, thereby suggesting that Biot model may not be the most appropriate theoretical framework for solving the inverse problem. The most recent results will be reviewed and future research directions will be highlighted to overcome the observed inconsistencies.

1:15

5pBB2. Structural and mechanical assessment of trabecular bone by *in vivo* magnetic resonance imaging. Felix Wehrli (Dept. of Radiology, 3400 Spruce St., Philadelphia, PA 19104, wehrli@mail.med.upenn.edu)

Bone is a complex composite material whose strength is determined by a combination of the material's intrinsic mechanical properties, its overall volume fraction, and architecture at the macro-, micro- and ultrastructural levels. Trabecular bone, the type of bone prevalent in the vertebrae and ends of long bones, is particularly prone to fracture in osteoporosis. It consists of a network of interconnected plates and struts of about 100-μm thickness immersed in a matrix of marrow. Advances in magnetic resonance imaging technology, including development of high-field magnets and radiofrequency coil technology, along with improved image processing and analysis techniques, now allow acquisition of images from which the three-dimensional (3D) microarchitecture can be retrieved, at least at peripheral skeletal locations such as the distal radius and tibia. From these data measures of topology, scale and orientation of the trabecular network, all contributing to the bone's mechanical competence, can be derived with high-serial reproducibility. The processed images can also be used to create 3-D meshes and mechanical parameters such as elastic and shear moduli computed by micro-finite-element analysis. The translation of this technology to clinical medicine is highlighted with examples from recent drug intervention studies, allowing direct assessment of structural and mechanical consequences of treatment.

1:30

5pBB3. The effect of structural anisotropy on the fast wave propagation in cancellous bone. Mami Matsukawa, Katsunori Mizuno (Lab. of Ultrasonic Electron, Doshisha Univ., Kyotanabe 610-0321, Japan, mmatsuka@mail.doshisha.ac.jp), and Yoshiki Nagatani (Kobe City College of Technol., Kobe, 651-2194, Japan)

QUS parameters are closely related to the structural properties and elastic properties of bone, which can provide important information related to bone quality and bone strength. One should be very careful, however, that bone contains complicated structure from microscopic to macroscopic levels. For example, the cancellous bone inside the epiphysis is composed of a complicated trabecular network in the bone marrow, showing strongly anisotropic and heterogeneous structure. We have then observed the longitudinal wave propagation in this complicated medium, paying special attention to the two wave phenomenon. The fast wave, which mainly propagates in the trabecular part of the cancellous bone, is more sensitive to the network structure. With the help of x-ray micro-computed tomography, three dimensional anisotropic trabecular structure and fast wave propagation were investigated using bovine cancellous bones. We can find the strong effect of anisotropic trabecular structure on the fast wave velocity. The mean intercept length (the information of average trabecular length) was a good parameter to describe the three dimensional trabecular structure and showed good correlation with the fast wave velocity.

Contributed Papers

1:45

5pBB4. Multi-emitters and multi-receivers probe for long cortical bone assessment. Jean-Gabriel Minonzio, Maryline Talmant, and Pascal Laugier (UPMC Univ Paris 06, UMR 7623, LIP, 15 rue de l école de médecine, F-75006 Paris, France)

Different quantitative ultrasound techniques are currently developed for clinical assessment of human bone status. This paper is dedicated to axial transmission: emitter(s) and receiver(s) are linearly arranged on the same side of the skeletal site. A multiparameter approach might be relevant to improve bone diagnosis and this be could achieved by accurate measurement of guided wave phase velocities. Clinical requirements and bone/soft tissue heterogeneities constrain the length probe to about 10 mm. Thus efficiency of conventional spatio-temporal Fourier transform is reduced. Signal processing to obtain reliable guided wave velocities is a key point. Here the guided mode phase velocities are obtained using a projection in the singular vectors basis determined by the singular values decomposition of the transmission matrix between the two arrays. This method has been first validated on metallic plates. A set of excised human radius were then probed. Attention was paid to observation of cut-off frequencies because of their high power of discrimination of relevant bone properties (thickness and elasticity). The analysis of the whole spectrum including several branches of intermediate phase velocities is under progress. The so-called bi-directional correction of the soft tissue thickness variations under the probe is used for the *in vivo* measurements.

2:00

5pBB5. Discrimination of bone fractures by low-frequency axial transmission velocity in the radius and tibia. Petro Moilanen (Dept. of Phys., Univ. of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland, petro.moilanen@jyu.fi), Mikko Määttä (Univ. of Oulu, 90014 Oulu, Finland), Vantte Kilappa, Leiting Xu (Univ. of Jyväskylä, 40014 Jyväskylä, Finland), Timo Jämsä (Univ. of Oulu, 90014 Oulu, Finland), Jussi Timonen, and Sulin Cheng (Univ. of Jyväskylä, 40014 Jyväskylä, Finland)

Low-frequency axial transmission velocity of the first arriving signal (V_{LF}, 0.4 MHz) has a wavelength (8–10 mm) sufficient to probe osteoporotic changes in subcortical bone. Purpose of the present study was to evaluate V_{LF} on retrospective fracture discrimination, using a custom-made ultrasonometer in the midshaft radius and tibia. Preliminary data for 49 nonfractured (NF; 45–87 years) and 16 fractured (F; 56–81 years) postmenopausal females were analyzed. The fractures included were caused by low or moderate trauma and mostly occurred in the forearms or lower legs. Subjects with disease or medication affecting bone metabolism were excluded. Areas under receiver operating characteristic curve were 0.76 for the radius and 0.60 for the tibia V_{LF}. When adjusted for age and BMI, odds ratio for the radius V_{LF} was 1.97 (95% CI: 1.11–3.48), while that for the tibia V_{LF} was not statistically significant. In the contrary, dual-energy x-ray absorptiometry (DXA; whole-body, femoral neck, and L2–L4) was unable to differentiate between the F and NF groups. Despite the small study population, these results suggest that V$\mathfrak{LF}$ in the radius discriminates osteoporotic fractures in cases of peripheral fractures, in particular, for which DXA has a limited sensitivity.

2:15

5pBB6. Decomposition of two-component pulses: Simulation and phantom experiment. Keith A Wear (FDA Ctr. for Devices and Radiological Health, Bldg. 62, Rm. 3108, 10993 New Hampshire Ave., Silver Spring, MD 20993)

Porous media often support propagation of two compressional waves. When cancellous bone samples are interrogated in through-transmission with broadband sources, these two waves often overlap in time. A method for measuring attenuation and velocity of the two component waves was developed. The method (1) assumes that the transfer function is the sum of two exponentially modulated sinusoids and (2) minimizes the sum of the squared error between a model fit and the data. The method was tested for decomposing a 500-kHz-center-frequency signal containing two overlapping components: one passing through a polycarbonate plate (fast wave) and another passing through a cancellous-bone-mimicking phantom (slow wave). The method yielded estimates of attenuation slopes accurate to within 7% (polycarbonate plate) and 2% (cancellous bone phantom). The method yielded estimates of phase velocities accurate to within 1.5% (both media). The method was tested on simulated data generated using attenuation slopes and phase velocities corresponding to bovine cancellous bone. Throughout broad ranges of signal-to-noise ratio and fast-slow-wave-velocity differential, the method yielded estimates of attenuation slope that were accurate to within 10% and estimates of phase velocity that were accurate to within 5% (fast wave) and 2% (slow wave).

2:30

5pBB7. Recovery of structural degradation indicators of human bones using elastic waves. Erick Ogam, Zine El Abeddine Fellah (Laboratoire de Mcanique et d'Acoustique, UPR 7051 CNRS, 31 chemin Joseph Aiguier, Marseille 13402, France), Jean-Philippe Groby (Universit du Maine, 72085 Le Mans Cedex 09, France), Robert Gilbert (Univ. of Delaware, Newark, DE 19716), and Armand Wirgin (Laboratoire de Mcanique et d'Acoustique, Marseille 13402, France)

Non-ionizing techniques of bone characterization have mostly been based on ultrasonic wave propagation. In this study, bone is characterized via its linear response to vibratory solicitations. The analysis of the response of bones to transient mechanical excitation is made both theoretically and experimentally. An orthotropic three-dimensional finite element interaction model (FEIM) using computerized tomography scan image geometries of excised dry human tibiae bones is developed. The model parameters (nine elastic constants) are recovered by solving an inverse eigenvalue problem using resonance frequencies from vibration spectroscopy. Transient waves are excited and signals acquired along the diaphysis of the tibiae, experimentally by employing piezoelectric transducers and numerically using FEIM and a time integration scheme. The elastic waves propagating as

guided modes in the bone are investigated using reassigned bilinear time-frequency distribution. To avoid ambiguity in determining the elastic wave phase velocities due to the dispersive nature of propagation in bone, the longitudinal $L(0,1)$ and flexural $F(1,1)$ phase velocities, measured at their asymptotes, i.e., the bar and Rayleigh velocities respectively, are proposed as indicators for assessing the strength of long bones. Their sensitivity to changes in bone state like the effect of cortical thickness variation due to aging and osteoporosis is examined.

2:45

5pBB8. On the relation between effective properties of cancellous bones and quantification of the microstructure. Miao-jung Yvonne Ou (Joint Inst. for Computational Sci., Oak Ridge Natl. Lab., 1 Bethel Valley Rd., Oak Ridge, TN 37831) and Robert Lipton (Louisiana State Univ., Baton Rouge, LA 70803)

The effective properties of a two-phase composite depends both on the constituent materials and the microstructure. Starting from David J. Bergman's work on the integral representation formula (IRF) approach for deriving bounds on dielectric properties of composites with isotropic constituents, followed by Golden & Papanicolaou's work on establishing the link between IRF and spectral theory, the IRF is on now a firm mathematical background for further generalization from isotropic constituents to general constituents. The key feature we are looking for is the separation of information on material contrast from that of microstructure in the IRF. This is especially important for (visco)elastic composites because their current available IRFs do not have this feature, even when constituents are isotropic. In this talk, we will present a recently derived IRF for general (visco)elastic composites with the above feature. The relation between IRF and the quantification of microstructure (as spectrum of a generalized Helmholtz projection operator) will be explained. If time permits, numerical results will be presented and the connection between the IRF and the fabric tensor introduced by Stephen C. Cowin will also be discussed.

3:00—3:30 Break

Invited Papers

3:30

5pBB9. Multiscale structure-functional modeling of lamellar bone. Kay Raum (Julius Wolff Institut, & Berlin-Brandenburg School for Regenerative Therapies, Charité Universittsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany, kay.raum@charite.de), Quentin Grimal (Universit Pierre et Marie Curie-Paris 6, F75006 Paris, France and CNRS, F75006 Paris, France), and Alf Gerisch (Universitt Darmstadt, Darmstadt, Germany)

Bone is a natural example of achieving a unique combination and variability of stiffness and strength. One of the striking features of bone tissue is the ability to adapt to variable loading conditions by multiple but well organized structural arrangements of mineralized collagen fibrils at several levels of hierarchical organization. A profound understanding of the structure-function relations in bone requires both experimental assessment of heterogeneous elastic and structural parameters and theoretical modeling of the elastic deformation behavior. A bottom-up approach for experimental assessment and numerical modeling of the hierarchical structure from the nanoscale to the macroscale will be presented. Experimental data are obtained by scanning acoustic microscopy between 50 MHz and 1.2 GHz and provide anisotropic elastic and structural information at the lamellar (nanoscale) and at the tissue matrix (microscale) level. These data are directly translated into a finite element mesh. By numerical deformation analyses, the homogenized elastic stiffness tensor of the next hierarchical levels (microscale to macroscale) is derived. At each level the numerical results are cross-validated by experimental data. It will be shown that local variations in elastic anisotropy within the femoral shaft are related to an inhomogeneous strain distribution resulting from external stresses by weight and muscle forces.

3:45

5pBB10. Role of absorption mechanisms for ultrasound attenuation in cancellous bone: Macroscopic modeling and experiment. Michal Pakula and Mariusz Kaczmarek (Inst. of Mech. and Comp. Sci., Kazimierz Wielki Univ., Kopernika 1, Bydgoszcz, Poland, michalp@ukw.edu.pl)

Evaluation of the relative contribution of physical mechanisms responsible for attenuation of ultrasonic wave in cancellous bone is one of the crucial issues from the point of view of modeling elastic wave propagation and related model-based identification of the structural and mechanical properties of bone material. Considering trabecular bone as a porous material filled with fluid, the wave attenuation may stem from (i) intrinsic absorption in the fluid and solid phase, (ii) friction at the fluid-solid interface, as well as (iii) wave scattering by inhomogeneities (pores/trabeculaes). The commonly used for modeling ultrasound propagation in cancellous bone of the macroscopic Biot's theory will be discussed in context of its potential applicability for prediction of wave parameters: phase velocity and attenuation coefficient as functions of frequency. Since the model was introduced for long wavelength range, the scattering effects are neglected, and the analysis will be focused on the absorption mechanisms responsible for attenuation of ultrasonic waves in bone material. The suitability of the model will be verified by comparison of results of sensitivity analysis of the model with *in vitro* experimental ultrasonic data obtained for cancellous bones filled with different fluids.

Contributed Papers

4:00

5pBB11. Characterization of cortical bone fracture with scanning confocal ultrasound and longitudinal acoustic velocity. Yi-Xian Qin, Jiqi Cheng, Suzanne Ferreri, and Wei Lin (Dept. of Biomedical Eng., Stony Brook Univ., 350 Psych.-A Bldg., Stony Brook, NY 11794, yi-xian.qin@sunysb.edu)

Non-invasive assessment of fracture, particularly in non-typical fracture, is a critical health problem. As a promising alternative to the x-ray, ultrasound has demonstrated potentials in early fracture diagnosis. A real-time scanning confocal ultrasound image was developed to evaluate bone defect and bone loss. The objectives of this study were to evaluate the cortical fracture gap size using quantitative ultrasound imaging and the longitudinal ultrasound velocity in bone to predict the fracture gap size. The total fractures were created by MTE compression at the middle diaphysis, with gap size varied from 1–5 mm ($N=4$). Fractures were tested with the ultrasound scanning with 0.5-mm resolution. The measured ultrasound signals were analyzed to calculate BUA (dB/MHz), ATT (dB), and transverse velocity (m/s). The longitudinal wave velocity was tested using three surface transducers. Strong correlations were observed between ultrasound and x-ray images in fracture size ($R2=0.91$). High correlation was found between gap size and the longitudinal velocity from 4000 to 3000-m/s for 0–5-mm gaps ($R2=0.93$). These results suggest that ultrasound is capable to predict bone fracture

and provide useful information for longitudinal assessment of complications, such as non-union fracture, and for evaluating healing. [Work supported by NSBRI/NASA and NIH.]

4:15

5pBB12. Dynamic acoustoelastic testing for noninvasive detection of microdamage in cortical tissue of long bones. Guillaume Renaud, Maryline Talmant, and Pascal Laugier (Laboratoire d'Imagerie Paramétrique, CNRS UMR7623, Université Pierre et Marie Curie, 15 rue de l'école de médecine, 75006 Paris, France, renaud_gu@yahoo.fr)

First developed for trabecular bone, dynamic acoustoelastic testing (DAT) is based on the measurement of ultrasound (US) pulses time of flight and energy modulations (TOFM and EMs) induced by a low-frequency (LF) acoustic wave, synchronously injected in the probed medium. The technique needs the LF wave to be quasi-uniform in the investigated region and quasi-static in time (compared with the US TOF). US TOFM and EM are related to elastic and dissipative nonlinearities, respectively. Here is presented the application of the technique to cortical tissue of long bones. A LF vibration is generated along the axis of the sample whose frequency corresponds to the first compressional resonance mode. Simultaneously, US pulses are emitted and received by a dedicated probe, after propagation along the axis of the sample. The emitter-receiver distance is chosen and the position of the US probe is adjusted so that the head wave arrives first in time. US TOFM and EM are thus computed using the head wave. After validation in an aluminum cylinder where no dissipative nonlinearity but weak quadratic elastic nonlinearity is observed, *in vitro* results are presented for bone. DAT is performed on diaphyseal bone sections before and after mechanical damage was induced.

4:30

5pBB13. Bone quality assessment at multiple skeletal sites using quantitative ultrasound imaging. Frederick Serra-Hsu, Jiqi Cheng, Wei Lin, and Yi-Xian Qin (Dept. of Biomed Eng., SUNY Stony Brook, HSC T18, Rm. 030, Stony Brook, NY 11794, fserrahs@ic.sunysb.edu)

Osteoporosis is responsible for nearly 1.5×10^6 fractures annually in U.S. with the most typical sites at the hip, spine, and wrist. Early diagnosis is essential for the prediction of fracture risk. While BMD only represents quantitative information of bone, current quantitative ultrasound (QUS) typically employs index using attenuation (BUA) and velocity (UV) measurements at the calcaneus. The objective of this study aims to develop an image based QUS technology for enhanced diagnostic readings at multiple anatomical sites. With a newly developed scanning confocal acoustic diagnosis system, five embalmed cadaver forearms were disarticulated from the wrist joint, cut at the mid shaft, leaving the interosseous membrane intact, and scanned by the ultrasound. Acoustic phantoms scanned with the new system and basic water coupling show low variation ($<7\%$ of mean BUA) with highly linear BUA curves (mean $R2=0.96$). BUA and UV were analyzed at the distal radius using existing algorithms; however, the distal radius geometry provides much different BUA and UV schemes than the calcaneous so site-specific calculation algorithms are needed. DXA, mCT, and mechanical testing are currently under evaluation. These preliminary data demonstrated feasibility of ultrasound imaging in critical skeletal site, i.e., wrist. [Work supported by NSBRI/NASA and NIH.]

4:45

5pBB14. Separating individual modes from multimodal guided wave signals in long cortical bones. K. L. Xu, D. A. Ta (Dept. of Electron. Eng, Fudan Univ., Shanghai 200433, China), P. Moilanen (Univ. of Jyväskylä, Finland), and W. Q. Wang (Fudan Univ., Shanghai 200433, China)

Time-frequency representation (TFR) is one of the classical methods used for guided wave signal processing in assessment of long cortical bones. In this paper, we first used a crazy-climber ridge detection method to separate individual modes from multimodal guided wave signals in the time-frequency domain. Second, a penalization algorithm, based on spline smoothing, was employed to reconstruct the individual modes separately in the time domain. Robustness of these methods was evaluated by analyzing highly noisy simulated signals and experimental signals on bovine tibiae. It was found that the separated TFR ridges, provided by the crazy-climber algorithm, were clearer than using the TFR alone and these corresponding trajectories were in good agreement with theoretical dispersion curves. The penalization algorithm, on the other hand, showed promising performance by yielding good accordance between the reconstructed individual modes and the theoretical predictions. Both the crazy-climber and the penalization algorithms were thus useful for TFR analysis of multimodal signals, even under strongly noisy conditions, by separating and reconstructing the individual contributions in the time domain. These methods can thus significantly improve ultrasonic guided wave signal processing by providing a feasible quantitative assessment of individual modes in temporal multimode signals, recorded in long bones.

Session 5pEA

Engineering Acoustics, Acoustical Oceanography, and Underwater Acoustics: Computer Modeling for Complex Acoustic Environments

Kenneth M. Walsh, Chair
K. M. Engineering, Ltd., 51 Bayberry Ln., Middletown, RI 02842

Chair's Introduction—1:00

Invited Papers

1:05

5pEA1. Stratified layer imaging using synthetic aperture processing. Steven Schock and Victoria Ringle (Dept. of Ocean and Mech. Eng., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431)

Synthetic aperture processing is applied to seabed reflection data to improve images of stratified sediment layering. Vertical profiles of sediments are limited in dynamic range by sediment scattering. Synthetic aperture processing, which increases the effectiveness along track length of the sonar transmitter and receiver, suppresses scattering noise, thereby allowing detection of stratified layering with weaker reflection coefficients. The image improvement resulting from the application of synthetic aperture processing is determined by comparing reflection profiles generated with and without synthetic aperture processing. The processing algorithm used in the analyses is not the conventional synthetic aperture processing algorithm which is coherent for point scatterers, but is an algorithm that provides coherent processing for planar reflectors.

1:25

5pEA2. Fast array processor for the noiselet reverberation model. Edmund J. Sullivan, Robert P. Goddard (Prometheus Inc., 21 Arnold Ave., Newport, RI 02840, ed@prometheus-us.com), Hyman A. Greenbaum (Information Technol., Middletown, RI 02842), and Carlos Godoy (Naval Undersea Warfare Ctr., Newport, RI 02841)

The noiselet reverberation model, instead of transmitting the pulse itself, transmits a superposition of many copies of the pulse, each copy with a random amplitude and phase. This allows the simulation to mimic the point scatterer method without requiring a prohibitively large number of scatterers. It uses the GRAB ray model to introduce the propagation information. It provides a real-time high-fidelity simulation, providing element-level time series as an output for a planar array of receivers. However, it is presently limited by the phase shift calculations required by the frequency-domain beamformer to about 20 receivers without an unacceptable trimming of the number of rays used. Here, an approach is demonstrated that significantly improves the speed of this computation by identifying those receivers requiring the same phase shift to within a given specified error, such that only a single phase shift computation is required. These receivers fall onto a finite width strip, or quasi stave, that is not required to be parallel to element rows or columns. Along with providing a significant speed-up to the calculation, the approach becomes more efficient as the number of receivers increases, thereby allowing real time simulations using receiver arrays of over 100 elements.

1:45

5pEA3. Multidomain modeling of transducers and acoustical systems. Stephen C. Thompson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

Multidomain modeling, recently also called multiphysics modeling, is appropriate for systems in which different parts of the system are modeled using different equations and different modeling variables. Each such domain is modeled using the variables and equations that are appropriate for it, and the domains are coupled by additional equations at their boundaries and in regions of overlap. Acoustic transducers are an obvious example of multidomain systems. Historically, electrical circuit analog modeling was an approach for multidomain analysis. In addition, specific domain couplings have been hard-coded into particular FEA codes for the analysis of some kinds of coupled systems. Today there are FEA codes, as well as codes based on simpler ordinary differential and algebraic equation analysis, that provide a robust framework for this kind of modeling. This paper will discuss these anaylsis methods and provide computational examples from some simple multidomain systems that include acoustic transducers. [Work supported in part by the Office of Naval Research.]

2:05

5pEA4. Ducted propagation in the atmosphere, from audible sound to infrasound. Roger Waxler, Carrick Talmadge, Kenneth Gilbert, Xiao Di, Phillip Blom, Jelle Assink, and Claus Hetzer (NCPA, Univ. of Mississippi, University, MS 38677)

Earth's atmosphere supports complicated and highly variable winds and temperature gradients. Temperature inversions and wind jets provide ducts in which sound can propagate efficiently to great distances from the source. Due to the intrinsic variability of the atmospheric winds, these sound ducts are also quite variable, both temporally and spatially. Further, the direction of the atmospheric winds

can change dramatically with increasing altitude so that the behavior of the propagated field can become quite complex as the range from the source increases. Despite this, some systematic behavior has been observed in long range sound propagation. An overview will be presented, with an emphasis on the systematic behavior. The current state of our understanding of the sources and consequences of the atmosphere's variability will be discussed.

Contributed Papers

2:25

5pEA5. Atmospheric sound scattering model to test signal coding methods for acoustic wind profiling. Paul Kendrick and Sabine von Hünerbein (School of Computing, Sci. and Eng., Univ. of Salford, Salford M5 4WT, United Kingdom, p.kendrick@salford.ac.uk)

This presentation concerns simulation of atmospheric sound propagation, including the backscatter of acoustic waves from moving temperature fluctuations. The model has been developed to evaluate the performance of acoustic wind speed profiling techniques, in particular, the use of pulse coding methods which aim to increase range resolution, accuracy, and data availability. Stepped frequency chirps are propagated through the model and the results used to optimize matched filtering algorithms. Results from simulations are presented and the performance of several different signal detection algorithms evaluated. This methodology provides insight into particular problems encountered by the signal detection algorithm and will be utilized to aid the design of an experimental acoustic wind speed profiler. [Work funded by the Engineering and Physical Sciences Research Council (EPSRC).]

2:40

5pEA6. A finite difference model for low frequency sound measurement in kraft recovery boilers. Robert Hildebrand (Lake Superior State Univ., 650 W. Easterday Ave., Sault Ste. Marie, MI 49783, rhildebrand@lssu.edu), Matthew Carroll (Texas A & M Univ. at Galveston, Galveston, TX 77553-1675), Ville Jarvinen, and Juha Miettinen (Tampere Univ. of Technol., Tampere, Finland)

Kraft recovery boilers greatly enhance the efficiency of paper pulp mills by burning the organic wastes recovered from the pulp making process and generating electricity, but significant smelt-water explosion hazards exist when steam is used as the working fluid due to sodium compounds in the sludge from the waste incineration process. Current practice is to use acoustic emission sensors to monitor for leaks in the bottom wall and to initiate an automatic shutdown sequence for the boiler once a leak is detected. A high-frequency model previously developed by the authors and verified by measurements performed in Skoghall, Sweden indicated strong attenuation in the sludge between the point of leakage and the detectors, but experimental results demonstrated a pass band at low frequencies (under 20 kHz). In this paper the authors develop a two-degree-of-freedom finite difference model for the prediction of acoustic attenuation at low frequencies. The bottom wall is modeled as a periodic tube and fin arrangement, and a time-domain finite difference method is used whereby the fin and tube are simple lumped elements. The model is then compared to previous experimental results, with which excellent agreement is obtained.

2:55

5pEA7. Analytical and finite element modeling of vibro-acoustic modes of a fluid-filled thick-walled hollow microcylinder and applications in low-power particle manipulation. Muhammet K. Araz and Amit Lal (SonicMEMS Lab., School of Elec. and Comput. Eng., Cornell Univ., Ithaca, NY 14853, mka22@cornell.edu)

Nonlinear acoustic interactions such as acoustic radiation forces or acoustic streaming require high-amplitude acoustic pressure or velocity terms to generate observable effects. While at the macro-scale these effects are not widely used; due to the highly compliant microstructures, high-amplitude acoustic pressure and velocity terms have been reached in a microscopic cavity even at very low-voltage drives (1–10 Vpp). Here, vibrational and acoustic modes of a thick-walled fluid-filled hollow cylindrical glass capillary are introduced. Vibrations are coupled to the capillary through a laser cut, C-shaped lead zirconate titanate plate. Mode shapes and dispersion relationships are obtained both analytically and from finite element modeling in ANSYS. Analytical modeling of the microcavity resonator shows excellent agreement with the computational models developed in ANSYS and the experimental results. Coupling of the vibrational modes of the capillary to the fluid enclosed plays an essential role in manipulation and separation of micro- and nanoparticles and biological entities in acoustically excited portable microfluidic platforms. In this regard, proper modeling of the vibroacoustic modes enables calculation of acoustic radiation force field generated inside the microfluidic capillary. Experimental results on particle collection and separation will be presented at the meeting.

3:10

5pEA8. Shock wave formation on turbulent coanda surfaces. Thomas Dowd (Dept. of Mathematics, James Madison Univ., 800 S Main St., Harrisonburg, VA 22807, dowd.thomas@gmail.com)

In the petroleum industry a procedure called flaring is used, which is the burning off of unwanted gas. Modern day flares use the Coanda effect,which states "when a jet of fluid is passed over a curved surface it bends to follow the surface entraining large amounts of air" to achieve smokeless combustion, increased combustion efficiency, and decreased thermal radiation due to the entrainment of large amounts of air. These advantages are at the cost of increased noise pollution, due to shock wave formation caused by the difference in the nozzle exit pressure and the ambient pressure. Sound called shock-associated noise is generated by the interaction of downstream propagating turbulent eddies and the stationary quasi-periodic shock-cell structure contained in the supersonic jet. A model of shock-associated noise near turbulent Coanda surfaces will be presented and compared with experimental data along with suggestions for reducing noise. Also a visualization of these shock wave formations and their intricacies will be presented if time allows.

3:25

5pEA9. Acoustic scattering from layered pentamode materials. Clyde L. Scandrett (Dept. of Appl. Mathematics, Naval Postgrad. School, Monterey, CA 93943), Jeffrey E. Boisvert, and Thomas R. Howarth (NAVSEA Newport, Newport, RI 02841-5047)

While receiving less attention in the literature than in the field of electromagnetic scattering, theoretical efforts to define and create acoustic skins capable of reducing scattered energy from obstacles by way of mimicking coordinate transformations through use of meta-materials have begun. The present work extends recent analysis of Norris by considering a variety of acoustic skins, from those comprised of fluid layers which are isotropic in bulk moduli with anisotropic density to those having anisotropic bulk moduli and isotropic density. In all but pure inertial types, fluid layers comprising the skins are pentamode materials governed by a special scalar acoustic equation for pseudo-pressure derived by Norris. In most cases presented, material properties of the fluid/pentamode layers are based upon target values specified by continuously varying properties resulting from theoretical coordinate transformations geared to minimize scattered pressure limited by realistic goals. The present work analyzes acoustic skins for the specific case of plane wave scattering from an acoustically hard sphere. An initial exploration of the parameter space defining such skins (for example, material properties of its constituent layers and operating frequency) is undertaken with a view toward "optimal" design. [Work supported by the NAVSEA Newport In-House Laboratory Individual Research (ILIR) Program.]

3:40

5pEA10. Power analysis for decoding of the digital audio encoding format MP3: Decoding the central processing unit and the graphics processing unit. Seung Gu Kang, Pil Joung Sun, Cheol Hong Kim (Chonnam Natl. Univ., Chonnam, Korea, chkim22@chonnam.ac.kr), and Jong-Myon Kim (Univ. of Ulsan, Ulsan, Korea)

Power consumption is one of the major design considerations for recent computing systems. MP3 players require MP3 decoding performed on the processor. We analyze the power consumption for MP3 decoding varying the processor type [central processing unit (CPU) and graphics processing unit (GPU)]. In general, the MP3 decoding is performed on CPU . GPU, which is a specialized processor that offloads three dimensional graphics rendering from the CPU, can be used for MP3 decoding since it is used for various computations as well as graphics rendering. At first, we analyze the power consumption for the MP3 decoding performed on CPU by using Inspector tool. And then, we analyze the power consumption for the MP3 decoding performed on GPU. To analyze the power efficiency of the GPU for MP3 decoding, we use CUDA for translating the MP3 decoding process from CPU to GPU. Computer unified device architecture (CUDA) is the computing engine in NVIDIA GPUs that is accessible to software developers through industry standard programming languages. According to our experimental results, the MP3 decoding performed on GPU saves power consumption by 23% compared to that on CPU. [Work supported by the MKE, Korea, under the ITRC supervised by the IITA (IITA-2009-(C1090-0903-0008).]

FRIDAY AFTERNOON, 23 APRIL 2010 GRAND BALLROOM V, 1:00 TO 3:15 P.M.

Session 5pPAa

Physical Acoustics: Outdoor Sound Propagation

D. Keith Wilson, Cochair
U.S. Army Cold Regions Research Lab., 72 Lyme Rd,, Hanover, NH 03755-1290

W. C. Kirkpatrick Alberts, Cochair
U.S. Army Research Lab., 2800 Powder Mill Rd., Adelphi, MD 20783

Contributed Papers

1:00

5pPAa1. Ground effects in time-domain simulations of outdoor sound propagation. Didier Dragna, Philippe Blanc-Benon (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France, didier.dragna@ec-lyon.fr), and Franck Poisson (SNCF, Direction de l'Innovation et de la Recherche, 75379 Paris Cedex 08, France)

Finite-difference, time-domain (FDTD) methods are precise and powerful tools to solve linearized Euler equations. Interaction between the acoustic waves with local wind or temperature gradients as well as a complex topography can be taken into account. Recently, a time-domain boundary condition has been proposed [Cotté *et al.*, AIAA J. **47**, 2391–2403 (2009)] and has been implemented in a FDTD solver using methods developed in the computational aeroacoustics community [Bogey and Bailly, J. Comput. Phys. **194**, 194–214 (2004)]. Propagation of an initial pulse over a distance of 500 m in a two-dimensional geometry is considered in a frequency band up to 1200 Hz. Surface waves which propagate close to and parallel to impedance grounds are exhibited. To validate the obtained results, a comparison is made in the time-domain with an analytical solution. Ground effects are discussed depending on the impedance of the ground. [This work was performed using HPC resources from GENCI-IDRIS (Grant 2009-022203).]

1:15

5pPAa2. Metamodel predictions for long range speech intelligibility using parabolic equation calculations. Amanda D. Hanford, Andrew R. Barnard, Michael L. Jonson, and Jay D. Martin (Appl. Res. Lab, The Penn State Univ., P.O. Box 30, State College, PA 16804)

The Crank–Nicholson parabolic equation (CNPE) method was used to calculate speech intelligibility for long range propagation of voice messages. How intelligible a voice message will be at long range is very sensitive to atmospheric conditions and parameters such as source height and background noise levels. In order to identify this sensitivity, four independent variables were identified and varied throughout the study. These independent variables are the atmospheric refraction index, which takes into account both temperature and wind gradients within the atmosphere, ambient temperature, source height, and background noise level. CNPE computations were performed for 31 optimally chosen cases where all four of the independent variables were random variables and were inputted into a metamodel. This allows for curve fitting and extrapolation of the existing data to determine intermediate results for any operation scenario defined within the bounds of the current study. Upon doing this, it was determined that the atmospheric refraction index is the most sensitive variable. In addition, there is virtually no dependence on ambient surface temperature.

1:30

5pPAa3. Wind velocity measurements with acoustic daylight. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., Boulder, CO 80305, oleg.godin@noaa.gov), Vladimir G. Irisov, and Mikhail I. Charnotskii (ZelTechnologies LLC and NOAA/Earth System Res. Lab., Boulder, CO 80305-3328)

Ambient acoustic noise in ocean and atmosphere provides acoustic illumination, which can be used, akin to daylight, to visualize objects and characterize the environment [Buckingham *et al.*, Nature(London) **356**, 327–329 (1992)]. It has been shown theoretically [O. A. Godin, Phys. Rev. Lett. **97**, 054301 (2006)] that deterministic travel times of acoustic waves propagating in opposite directions between two points in an inhomogeneous, moving, time-independent medium can be retrieved from the cross-correlation function of non-diffuse acoustic noise recorded at the two points. Thus, the noise cross-correlation function contains information on non-reciprocity of sound propagation, which can be utilized to measure flow velocities small compared to the sound speed. This paper presents an experimental verification of the theoretical predictions. Deterministic travel times have been retrieved from cross-correlations of noise recorded at three microphones about 20 m apart and inverted for the sound speed and two horizontal components of the wind velocity. Traffic noise served as the acoustic daylight in the experiment. Accuracy of the passive acoustic measurements of sound speed and wind velocity has been confirmed by comparison with contact measurements of wind, air temperature, and humidity. Possible uses of acoustic daylight for a low-cost environmental monitoring will be discussed.

1:45

5pPAa4. Acoustic location under metal foam windscreens. Jeremy Webster, Richard Raspet (Nat. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677), Jason Beam (Birmingham-Southern College, Birmingham, AL 35254), and Kelsey Dudziak (Rhodes College, Memphis, TN 38112)

Large wind noise reduction outdoors can be achieved by the use of metal foam layers mounted flush with the surface. In this research, a triangular array for source direction determination is mounted under different thickness metal foams. The performance of the array in locating a spark source is determined by measurements with and without metal foam covers. The performance is quantified by agreement with the actual direction and by the chi squared values of the fit to the direction and the speed of sound. Some geometries produced large errors in speed of sound but the corresponding directions were reasonably accurate. Relatively thin layers of metal foam produced accurate estimates for the direction and sound speed. With the microphones mounted just under a 0.5-in.-thick foam, the direction could be measured to within a few degrees. [This work is funded by the Armaments Research, Development, and Engineering Command (ARDEC)].

2:00

5pPAa5. Gunshot localization using distributed acoustic sensors. Xiao Di, W. Garth Frazier, and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

For detecting and localizing gunshot sound, distributed acoustic sensors with various baseline separations (several meters to dozens of meters) are used in our field experiments. The large baselines of the distributed sensor networks, based on the time difference of arrival method, provide improved localization accuracy and have distinct advantages over small arrays. Results are reported and compared for various baseline separations, detection distances, numbers of sensors, and weather conditions. The sound source used is a propane cannon. Several least squares methods are used in the data analysis and source localization. Results from the experiment are presented and discussed. The localization errors are analyzed as a function of combined error sources (GPS, timing, weather, etc.). Possible ways to improve the localization accuracy are suggested. [Research supported by the U. S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

2:15

5pPAa6. Investigation of dominant sources of infrasonic wind noise in ground level microphones. John Paul R. Abbott, Richard Raspet, Jeremy Webster, and Jiao Yu (Dept. of Phys. and Astronomy, Natl. Ctr. for Physical Acoust, Univ. of Mississippi, University, MS 38677)

A two part project has investigated whether the pressure fluctuations (infrasonic wind noise) at the ground, as measured in an open field, can be predicted from the vertical wind velocity measured above the ground. The dominant region contributing to the pressure on the ground is calculated theoretically and displayed with contour plots. Experimentally, time dependent correlations between the measured pressure and wind velocities were done. The correlation coefficients were very low, typically between 0.08–0.15, for broadband frequencies up to 5 Hz. Cross spectrum analysis and application of octave band pass filters showed the dominant frequencies for correlations to be around 0.07–0.1 Hz, with correlation coefficients typically between 0.3–0.5. [Project funded by the US Army Space and Missile Defense Command.]

2:30

5pPAa7. Acoustic response of rigid pipe wind-noise reducers for infrasound measurement. Matthew E. Poese and Thomas B. Gabrielson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

A network of rigid pipes with multiple inlets is one commonly employed strategy to filter out pressure fluctuations generated by wind when making infrasound measurements. This talk presents a model of the transfer function of such a rosette filter using a computer model (DeltaEC) that includes thermo-viscous effects at the walls of the pipes. Measurements of this transfer function have been made utilizing the coherence between a calibrated reference microphone and a calibrated microphone shielded by the wind-reduction rosette and these data will be compared to the model. Rosettes with pipes of varied lengths were tested and modeled. The response of the rosette as a function of the vertical arrival angle of the incident plane wave will also be presented. [Work supported by US Army Space and Missile Defense Command.]

2:45

5pPAa8. Acoustical characterization of rigid porous materials. Kai Ming Li and Sheng Liu (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

In an earlier study [K. M. Li, J. Acoust. Soc. Am. Soc. **123**, 1352–1363 (2008)], an analytical expression has been derived to predict the penetration of sound into a rigid porous ground due to a monopole source. In the present study, this analytical formula has been exploited for a numerical method to characterize the acoustical properties of rigid porous materials. They can be obtained from a pair of excess attenuation spectra measured above and below a thick layer of the material. Specifically, its propagation constant and complex density ratio can be determined from the measured data by a downhill simplex method. A series of preliminary indoor measurements was conducted in an anechoic chamber for measuring the propagation of sound above and below a layer of glass beads of different sizes. Based on the measured excess attenuation spectra, the acoustical characteristics of two different sizes of glass beads were determined numerically. These acoustical parameters were compared with the predictions based on the classical impedance model.

3:00

5pPAa9. A theoretical and experimental study of sound propagation above a rigid porous ground. Sheng Liu and Kai Ming Li (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031, mmkmli@purdue.edu)

The sound propagation due to a monopole source near a rigid porous ground is reviewed. Closed form approximation solutions for predicting the sound field above the ground surface were derived analytically by (1) the method of pole subtraction and (2) the use of an effective impedance approach. The locations of the surface wave pole according to both methods were examined. These two approximate solutions were compared with an exact evaluation of the Fourier integral representing the sound field above the rigid porous ground. It is found that the approximate solution using the effective impedance approach offers relatively better agreements with the exact numerical solution. In addition, the effective impedance approach generally leads to a more accurate uniform asymptotic solution for the sound field above the rigid porous ground. In this presentation, the reason for this good agreement is explained. The finding is further supported by experimental data obtained by conducting indoor measurements over a layer of glass beads.

Session 5pPAb

Physical Acoustics: Ultrasonics

Andrea Prosperetti, Cochair
Johns Hopkins Univ., Dept. of Mechanical Engineering, 3400 N. Charles St., Baltimore, MD 21218

Andrew A. Piacsek, Cochair
Central Washington Univ., Dept. of Physics, 400 E. University Way, Ellensburg, WA 98926-7422

Contributed Papers

1:00

5pPAb1. Interaction of shear horizontal ultrasonic waves with a surface grating: Analytical and experimental results. Tony Valier-Brasier (Lab. d'Acoustique de l'Univ. du Maine (LAUM), UMR CNRS 6613, Ave. Olivier Messiaen, 72000 Le Mans Cedex 09, France, tony.valier-brasier.etu@univ-lemans.fr), Damien Leduc (Universit du Havre, 76610 Le Havre, France), Catherine Potel (Univ. du Maine, 72000 Le Mans Cedex 09, France), Bruno Morvan (Univ. du Havre, 76610 Le Havre, France), Michel Bruneau (Univ. du Maine, 72000 Le Mans Cedex 09, France), and Jean-Louis Izbicki (Univ. du Havre, 76610 Le Havre, France)

This paper deals with the propagation of shear horizontal (SH) ultrasonic waves on a plate with one surface grating. The theoretical acoustic field in the corrugated plate is obtained from a linear combination of the Neumann eigenmodes of a plane plate waveguide which bounds outwardly the perturbed surface of the plate considered. Coupling of the eigenmodes with the rough walls is studied using the integral formulation. The effect of the roughness is expressed in such a way that two intermodal coupling mechanisms are highlighted: a bulk coupling and a surface coupling. The first one depends only on the depth of the roughness and the second one on both its depth and its slope. Moreover, the coupling between SH modes is studied experimentally. An electromagnetic acoustic transducer enables a noncontact emission of the SH waves. Notably, the energy transfer from a mode to others so appears to be strongest at frequencies where a relation between the period of the grating and the wavenumbers of the coupled SH modes is verified.

1:15

5pPAb2. High-quality acoustic cavities based on radial sonic crystals. Daniel Torrent (Dept. Electron Eng., Polytechnic Univ. of Valencia, datorma1@upvnet.upv.es) and José Snchez-Dehesa (Polytechnic Univ. of Valencia.)

Radial sonic crystals are radially inhomogeneous systems that present a dispersion relation similar to one dimensional sonic crystals, with the peculiarity of having low-frequency band gaps. These structures are here employed to design high-quality circularly symmetric acoustic cavities. The presence of band gaps in their dispersion relation allows to localize sound in the cavity centers, leading to strong resonances whose frequency depends on the symmetry of the field. Therefore, it is possible to design cavities with monopolar, dipolar, quadrupolar, etc., symmetry. A proposal for its physical realization is also described, showing that radial sonic crystals are possible to build by using the concept of acoustic metamaterials. [Work supported by MCIIN of Spain and ONR.]

1:30

5pPAb3. Ultrasonic characterization of fluid saturated absorbing polymethylacrylate porous plates. Serge Derible (LOMC GOA, FRE CNRS 3102, FANO FR CNRS 3110, Univ. Le Havre, 76600 Le Havre, France), Pierre Campistron, Georges Freiha (Univ. de Valenciennes et du Hainaut Cambrésis, 59313 Valenciennes, France.), Hervé Franklin (Univ. Le Havre, 76600 Le Havre, France), and Bertrand Nongaillard (Univ. Valenciennes et du Hainaut Cambrésis, 59313 Valenciennes, France)

Experiments are led on five manufactured porous polymethylmethacrylate (PMMA) samples, numbered 1–5. Their constituting materials are packed beads whose diameters run from 200–300 μm for plate 1 up to 600–700 μm for plate 5. The plates are normally insonified by wide band transducers acting within the frequency range 100–800 kHz. The samples obey Biot's theory taking into account the viscoelasticity of the solid frame. The reflection and transmission coefficients are measured and then added or subtracted to give the two transition terms, issued from S-matrix theory. They are connected to the symmetrical or asymmetrical motions of the faces of the plates. Those terms allow us a characterization of the porous media by means of a computer program that seeks the values of the Biot parameters giving the best agreement between theory and experiments.

1:45

5pPAb4. Evaluation of surface waves in bone-conducted sound. Margaret Wismer, William O'Brien (BRL, Univ. of Illinois, Urbana, IL), Odile Clavier, and Anthony Dietz (Creare, Hanover, NH)

In order to better understand the basic physical principles of bone-conducted sound, the skull is approximated as a thin spherical fluid-filled shell. Results for sound waves propagating through and around this structure have been determined for a range of frequencies, media, and shell sizes. Since the early 1950s researchers have studied the problem of what happens when waves hit solid targets. Most of the research is concerned with the back-scatter signal but recognize that the interaction of the wave with the object influences these results. For the bone-conducted sound study, the interaction between the incident signal and the object is the critical information. For acoustic noise interacting with the human skull, the ka is below 5. The structure supports the existence of surface waves as the predominant mechanism of energy transfer within the skull. Multiple types of surface wave for this structure of a thin, curved fluid-solid interface exist. Signals which propagate on the air side of the surface due to the curvature are known as Franz or Stoneley waves. Since the shell is thin compared to the typical wavelength of sound, waves within the shell are lower order Lamb or plate waves. All of these waves are dispersive in nature with group speeds considerably smaller than the speed of sound in bone.

2:00

5pPAb5. Acousto-optic monitoring of high-intensity focused ultrasound lesion formation in optically diffuse tissue. Puxiang Lai, James R. McLaughlan, Andrew B. Draudt (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), Todd W. Murray (Univ. of Colorado at Boulder, Boulder, CO 80309), Robin O. Cleveland, and Ronald A. Roy (Boston Univ., Boston, MA 02215)

Tissue heating by high-intensity focused ultrasound (HIFU) is a promising modality for minimally invasive therapy. However, real-time treatment monitoring still poses significant challenges, particularly at the lower exposure levels where stable cavitation and/or boiling does not result. Bubble-

-free HIFU lesions offer little acoustic contrast; however, one does observe significant contrast in both optical scattering and absorption. We employ acousto-optic (AO) imaging to sense, in real time, optical changes induced by lesion formation. By using a transducer to simultaneously heat a tissue volume and pump the AO interaction, lesions generated in excised chicken breast are monitored in real time. The change in AO response with time is linearly related to the time-dependent lesion volume, provided the diameter of the lesion does not exceed the width of the optical beam. Therefore, AO sensing can be used to both determine the onset of lesion formation and the resulting volume of the necrosed region. The feasibility of using the observed change AO signal amplitude as the criteria to guide HIFU exposure in real time is demonstrated. [Work supported by the Center for Subsurface Sensing and Imaging Systems, NSF ERC Award No. EEC-9986821.]

2:15

5pPAb6. Fiber based transduction. Ethan N. Phung, Barbara Deschamp, and Charles Thompson (Dept. of Elec. and Computing Eng., Univ. of Massachusetts Lowell, One Univ. Ave., Lowell, MA 01854, narinphung86@hotmail.com)

In this work we will examine the conversion of mechanical vibration into electrical energy. This is done by using a conductive fiber in close proximity to a conductive plate. In this configuration it is shown that the fiber behaves as a square law detector. This observation may make it useful for mechanical power conversion for radio frequency demodulation. Based on the previous findings, the mechanical vibration of the fiber can be modeled by a simple damped oscillator with external applied source. The result indicated that the original signal can be recovered.

FRIDAY AFTERNOON, 23 APRIL 2010 DOVER A/B/C, 1:00 TO 3:20 P.M.

Session 5pSC

Speech Communication: Machine Learning Techniques for Speech Recognition

Mark A. Hasegawa-Johnson, Chair
Univ. of Illinois, ECE, 405 N. Matthews, Urbana, IL 61801

Chair's Introduction—1:00

Invited Papers

1:05

5pSC1. A novel algorithm for sparse classification. Sujeeth Bharadwaj and Mark Hasegawa-Johnson (Beckman Inst., Univ. of Illinois at Urbana-Champaign, sbhara3@illinois.edu)

A recent result in compressed sensing (CS) makes it possible to perform non-parametric speech recognition that is robust to noise and that requires few training examples. By taking fixed length representations of training examples and stacking them in a matrix, a frame or an over-complete basis can be constructed. Gemmeke and Cranen showed that sparse projections onto this frame recover the correct transcription with 91% accuracy at -5-dB SNR. The goal of speech recognition is not sparse projection onto training tokens, but onto training types. Sparse projection onto types can be achieved by building a frame for each word in the dictionary and stacking the frames to form a rank 3 tensor. Speech recognition is performed by convex linear projection onto the tensor, with sparsity enforced only in the index that specifies type. A mixed L1/L2 relaxation was derived, leading to a Newton descent algorithm that converges to the global optimum.

1:25

5pSC2. Speech classification using penalized logistic regression with hidden Markov model log-likelihood regressors. Øystein Birkenes (TANDBERG, Philip Pedersens vei 20, 1366 Lysaker, Norway, oystein.birkenes@tandberg.com), Tomoko Matsui (The Inst. of Statistical Mathematics, Tokyo, Japan), Kunio Tanabe (Waseda Univ., Tokyo, Japan), and Magne Hallstein Johnsen (Norwegian Univ. of Sci. and Technol., Trondheim, Norway)

Penalized logistic regression (PLR) is a well-founded discriminative classifier with long roots in the history of statistics. Speech classification with PLR is possible with an appropriate choice of map from the space of feature vector sequences into the Euclidean space. In this talk, one such map is presented, namely, the one that maps into vectors consisting of log-likelihoods computed from a set of hidden Markov models (HMMs). The use of this map in PLR leads to a powerful discriminative classifier that naturally handles the sequential data arising in speech classification. In the training phase, the HMM parameters and the regression parameters are jointly estimated by maximizing a penalized likelihood. The proposed approach is shown to be a generalization of conditional maximum likelihood (CML) and maximum mutual information (MMI) estimation for speech classification, leading to more flexible decision boundaries and higher classification accuracy. The posterior probabilities resulting from classification with PLR allow for continuous speech recognition via N-best or lattice rescoring.

1:45

5pSC3. Integrating evidence over time: A look at conditional models for speech processing. Eric Fosler-Lussier, Jeremy Morris, Ilana Heintz, and Rohit Prabhavalkar (Dept. of Comput. Sci. and Eng., The Ohio State Univ., 2015 Neil Ave., Columbus, OH 43210, fosler@cse-ohio-state.edu)

Many acoustic events, particularly those associated with speech events, can be thought of as events in a rich descriptive subspace where the dimensions of the subspace can be thought of as a sort of decomposition of the original event space. In phonetic terms, we can think of how phonological features can be integrated to determine phonetic identity; for auditory scene analysis we can look how features like harmonic energy and cross-channel correlation come together to determine whether a particular frequency corresponds to target speech versus background noise. Some success has been achieved by thinking of these problems as probabilistic detection of acoustic (sub-)events. However, event detectors are typically local in nature and need to be smoothed out by looking at neighboring events in time. This talk describes using conditional random fields models within the automatic speech recognition setting to combine bottom-up speech event detectors. The talk will explore some of the successes and limitations of this log-linear method which integrates local evidence over time sequences.

2:05

5pSC4. Posterior-based attributes in machine recognition of speech. Hynek Hermansky (Ctr. for Lang. and Speech Processing, The Johns Hopkins Univ., Baltimore, MD)

Current ASR uses two main sources of information, the top-down prior knowledge in the form of a statistical language and the bottom-up information from acoustic speech signal. The information from acoustic signal is presented in so called speech attributes that are derived from the signal. Any information that gets lost during attribute extraction cannot be recovered. Any preserved information that is irrelevant to the ASR task could complicate further processing. Attributes in conventional ASR are derived by some transformation of the short-term spectrum of speech that represents magnitude frequency components of a short segment of the signal. Origins of this representation can be traced to speech vocoding. For a past decade, we are working on attributes that are derived from posterior probabilities of speech sounds that are estimated by multilayer perceptron trained on large amounts of labeled data. This speech representation is now used in most leading state-of-the-art experimental systems. The talk will describe several research issues that influence the effectiveness of these posterior-based attributes. These issues include the choice of classes of speech sounds, input evidence for the posterior estimation, architecture and training of the posterior probability estimator, and post-processing of posterior estimates for use with conventional GMM-HMM ASR.

2:25

5pSC5. Machine learning strategies for recovering speech articulatory trajectories and gestures from speech. Vikramjit Mitra, Hosung Nam, Carol Y. Espy-Wilson, Elliot Saltzman, and Louis Goldstein (Dept. of Elec. and Comput. Eng., Speech Commun. Lab., Inst. of Systems Res., Univ. of Maryland, College Park, MD 20742, vmitra@umd.edu)

Articulatory information can improve the performance of automatic speech recognition systems. Unfortunately, since such information is not directly observable, it must be estimated from the acoustic signal using speech-inversion techniques. Here, we first compare five different machine learning techniques for inverting the speech acoustics generated using the Haskins Laboratories speech production model in combination with HLsyn. In particular, we compare the accuracies of estimating two forms of articulatory information (a) vocal tract constriction trajectories and (b) articulatory flesh-point pellet trajectories. We show that tract variable estimation can be performed more accurately than pellet estimation. Second, we also show that estimated tract variables can improve the performance of an autoregressive neural network model for recognizing speech gestures. We compare gesture recognition accuracy for three different input conditions: (1) generated acoustic signal and estimated tract variables, (2) acoustic signal and the original (or groundtruth) tract variables, and (3) acoustic signal only. Results show that gesture recognition accuracy was, not surprisingly, best for condition (2) and worst for condition (3). Importantly, however, condition (1) yielded better performance than (3), demonstrating that estimated tract-variable articulatory information is indeed helpful for automatic speech recognition.

2:45

5pSC6. Margin based discriminative training techniques for automatic speech recognition. Fei Sha (Dept. of Comput. Sci., Univ. of Southern California, Los Angeles, CA 90089), Chih-Chieh Cheng, and Lawrence Saul (Univ. of California, La Jolla, CA 92093)

Recently, large margin techniques have gained popularity for discriminative training of continuous-density hidden Markov models. These techniques are motivated by statistical learning theory, aiming to strike a balance between model complexity and generalization error. When used in speech recognition, these techniques separate discriminant scores from the correct transcriptions from all incorrect recognizer outputs. The parameters of acoustic models are adjusted to maximize the separation—margin—that is proportional to recognition errors. Many of such techniques have been successfully applied to large vocabulary speech recognition tasks and have improved the performance of existing systems with traditional approaches for parameter estimation. In this talk, I will review some of the work that my collaborators and I have undertaken in this direction. I will also describe a few new advances, including online learning of large margin separation as well as applying margin based techniques to adaptively transform acoustic features jointly with acoustic models.

Contributed Paper

3:05

5pSC7. Spoken Hindi paired word recognition using probabilistic neural network. Dinesh Kumar Rajoriya , R. S. Anand, and R. P. Maheshwari (Dept. of Elec. Eng., Indian Inst. of Technol. Roorkee, Roorkee, Uttarakhand 247 667, India, dineshrajoriya@gmail.com)

Automatic speech recognition system plays significant role in the development of real world applications. This paper demonstrated spoken Hindi (Indian national language) paired word recognition (SHPWR) system, which has been examined with the help of probabilistic neural network as a classifier. This type of network is a combination of radial basis layer and a competitive transfer function layer which picks the maximum probabilities as a final result. SHPWR is pattern recognition (PR) problem. A PR technique encompasses two fundamental tasks; description and classification. In the description process, features are extracted from the Hindi paired word templates, and probabilistic neural network is used as a classifier. For the experimental point of view 1000 Hindi paired word templates have been recorded from individuals with different environment condition, gender, and age groups. Speaker dependent SHPWR systems achieved close to 100% while the accuracy of speaker independent SHPWR systems is quite adequate.

FRIDAY AFTERNOON, 23 APRIL 2010 ESSEX A/B/C, 1:00 TO 3:35 P.M.

Session 5pSP

Signal Processing in Acoustics, Underwater Acoustics, and Animal Bioacoustics: Classification Methods in Acoustics and Non-Gaussian Noise II

R. Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Brett Bissinger, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Invited Papers

1:00

5pSP1. Formulation for statistics of echoes due to a finite number of scatterers and patches of scatterers in a directional sonar beam. Timothy K. Stanton (Dept. Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., MS 11, Woods Hole, MA 02543, tstanton@whoi.edu)

A general theoretical formulation is developed for echo statistics associated with a finite number of scatterers and patches of scatterers, each being randomly located in a directional sonar beam. This is the case of a direct-path sonar where there are no echoes from the seafloor and sea surface. The formulation is derived by combining the equations of Ehrenberg and colleagues (single scatterer randomly located in sonar beam) and Barakat (sum of finite number of random variables). Each scatterer or patch of scatterers is independent and can have its own echo probability density function (PDF) (before beampattern effects). Examples show that the PDF of the echo (as received by the sonar) is strongly non-Rayleigh due to beampattern effects and finite number of scatterers or patches. Numerical simulations are made for validation. [Work supported by ONR.]

1:20

5pSP2. Featureless classification for active sonar systems. Mary E. Soules and Joshua B. Broadwater (The Johns Hopkins Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723)

Active sonar systems depend on classification algorithms to identify target echoes and suppress false alarms. Historically, classifiers use a set of empirically derived features that exhibit some statistical separation between background clutter and target echoes. In an ideal scenario, these features would form a sufficient set of statistics capturing all of the information required to classify an echo return. Unfortunately, due to their empirically derived nature, features are rarely provably sufficient. To overcome this drawback, a featureless classifier will be presented. Instead of features, the raw data samples, which form a trivial but provable set of sufficient statistics, are used. The adaptive cosine estimate algorithm has a history of success with featureless classification in other applications and is well suited for underwater acoustics. An in-depth look at the featureless algorithm that will include comparisons to several traditional active sonar classifiers will be provided.

1:40

5pSP3. Clustering classification methods for acoustic applications including the analysis of clicking whale data. Juliette W. Ioup, George E. Ioup, Lisa A. Pflug (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, jioup@uno.edu), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA), Christopher O. Tiemann (Univ. of Texas at Austin, Austin, TX), and Charles H. Thompson (NOAA/NMFS/SEFSC, Stennis Space Ctr., MS)

Several clustering methods are available to group like objects into classes. The K means approach is well established. Both standard and homegrown algorithms have been tested in the following applications. One neural network technique is self-organizing map clus-

tering [T. Kohonen, *Self-Organizing Maps*, 2nd ed. Springer, New York (1997)]. Although it has some similarities to K means, it has notable differences which can be quite helpful in classification. An advantage is that it has the ability to limit the number of clusters automatically. This can be an important aspect when the number of classes is unknown. After a review of the basics of these techniques, the power of the methods is illustrated by applications to the study of clicking whales. The Littoral Acoustic Demonstration Center has acquired underwater acoustic data from clicking whales in the Gulf of Mexico, and it has access to data from the third and fourth International Workshops on Detection, Classification and Localization of Marine Mammals Using Passive Acoustics. Application of K means and SOM clustering to these data is an important component of identifying individual whales acoustically. Additional details on these results are being presented in Animal Bioacoustics sessions at this meeting. [Research supported by ONR and SPAWAR.]

2:00

5pSP4. Waveguide invariant minimum variance scatterer depth classification for active sonar. Ryan Goldhahn, Granger Hickman, and Jeffrey Krolik (Dept. of ECE, Duke Univ., 130 Hudson Hall, Durham, NC 27708-0291)

Active sonar systems are plagued by false alarms due to confusion between returns from water-column targets and backscatter from the bottom. Both feature-based and physics-based classifiers are notoriously susceptible to mismatch of the environment used for training and/or modeling active sonar returns. In this paper, in order to achieve more robust classification, uniformly sub-sampled DFT-coefficients from a single snapshot of the wideband active sonar return are used to define a waveguide-invariant spectral density matrix (WI-SDM). The WI-SDM facilitates adaptive matched-filtering based approaches for target depth estimation, where the waveguide invariant property is exploited to obtain uncorrelated snapshots without inflating covariance matrix rank. Depth classification is then performed by designing a waveguide-invariant minimum variance filter (WI-MVF) with adaptive weights which minimize ambiguous depth sidelobes. Simulation and real data results in a shallow-water Mediterranean environment are presented to illustrate the approach. [Work sponsored by ONR.]

Contributed Papers

2:20

5pSP5. Testing the temporal robustness of an automatic aural classifier. Stefan Murphy and Paul C. Hines (Defence R&D Canada, P.O. Box, 1012 Dartmouth, Nova Scofia, B2Y 3Z7, Canada stefan.murphy@drdc-rddc.gc.ca)

Military sonar systems must detect and classify submarine threats at ranges safely outside their circle of attack. However, in littoral environments, echoes from geological features (clutter) are frequently mistaken for targets of interest, resulting in degraded performance. Perceptual signal features similar to those employed in the human auditory system can be used to automatically discriminate between target and clutter echoes, thereby improving sonar performance. [J. Acoust. Soc. Am. 122, 1502–1517 (2007)] The present work examines the temporal robustness of the aural classifier using data from two field trials: the first in 2007 and the second in 2009. The experiments were conducted on the Malta Plateau using a cardioid towed-array receiver, and a broadband source transmitting linear FM sweeps from 600–3500 Hz. The data set consists of hundreds of pulse-compressed echoes from several surrogate targets and geological clutter objects. The echoes are examined using an automatic classifier that processes each echo to extract perceptual features. Each echo is classified as target or clutter based on the position vector formed by these features. The classifier establishes a boundary between clutter and target echoes in the feature space using the 2007 experiment. Temporal robustness is investigated by testing the classifier on echoes from the 2009 experiment. In this work, the experiments are reviewed and initial results are presented.

2:35

5pSP6. On the use of instantaneous acoustic intensity for the classification of defects in air-filled pipes. Kirill Horoshenkov, Tareq Bin Ali, Simon Tait (School of Eng., Univ. of Bradford, Bradford BD7 1DP, United Kingdom), and Alexandra Tolstoy (ATolstoy Sci., McLean, VA 22101)

The sound pressure in an air-filled pipe has been measured using six pairs of matched microphones. These data have then been analyzed to determine the six vectors of instantaneous particle velocity and acoustic intensity. The acoustic intensity vectors have been processed coherently using the matched field processing (MFP) technique to study the effect of changes such as blockages and cracks on the ambiguity surface. The results show that the technique is sensitive to very small changes in the pipe cross-section and the boundary conditions on the pipe wall. It enables us to locate a defect to within 1% of the measurement range. It also enables us to discriminate between blockages and wall cracks using the spectral characteristics of the ambiguity surface predicted via the MFP. This technique does not require acoustic modeling of the sound field in the pipe, but it relies on the sound pressure data recorded on three or more microphones in the absence and presence of a defect.

2:50

5pSP7. Source depth classification of passive sonar signals using amplitude statistics. R. Lee Culver, Colin W. Jemmott (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804), Brett E. Bissinger, and N. K. Bose (Penn State Univ., State College, PA 16804)

A number of at-sea measurements have shown that fluctuations in the amplitude of passive sonar signals are affected by source depth, range, and the propagation path from the source to the receiver. However, to date, received signal statistics have not been incorporated into passive sonar classification algorithms because the field results have not been accompanied by development of a statistics-based classifier architecture whose performance can be assessed using accepted signal processing metrics, e.g., a receiver operating characteristic curve. The performance of a statistics-based classifier depends critically upon statistical knowledge of the received signal, which in turn depends upon statistical knowledge of relevant environment parameters (sound speed in the water column and sediment, for example). Thus the classifier architecture must make use of the environmental information which in reality is only approximately known and must be described in statistical terms. Given a statistical description of the environment, an ocean acoustic propagation model and Monte Carlo simulation can be used to predict received signal statistics. Once received signal statistics are available for the different classes, a classifier can be designed. Here we examine a likelihood ratio binary source depth classifier for passive sonar. [Work sponsored by ONR Undersea Signal Processing.]

3:05

5pSP8. Minimum Hellinger distance classification for passive sonar in the presence of noise. Brett E. Bissinger (Dept. of Elec. Eng., and Appl. Res. Lab, Penn State Univ., P.O. Box 30, State College, PA 16804), R. Lee Culver, and N. K. Bose (Penn State Univ., State College, PA, 16804)

Passive source classification in the underwater environment is a challenging problem in part because propagation through the space- and time-varying media introduces variability and uncertainty in the signal. Acoustic

propagation models can predict received fields accurately, but they are sensitive to input environmental parameters which cannot be known exactly. This uncertainty in environmental knowledge used in signal predictions results in imperfect statistical class models. Classifiers that rely on simulations of the environment must therefore be robust to these imperfections. The minimum Hellinger distance classifier (MHDC) has been shown to be robust to such mismatches in situations with no noise. Low-noise situations allow demonstration of the properties of the classifier; however, real passive sonar tends to have significant noise levels. Therefore the MHDC's performance is examined and compared to that of a log-likelihood ratio classifier when applied to narrowband passive underwater acoustic signals in the presence of noise. Both classifiers are applied to synthetic Gaussian data, synthetic acoustic data, and actual acoustic data, all with noise. Classifiers are evaluated using receiver operating characteristic curves, a traditional performance metric in signal processing. [Work supported by ONR Undersea Signal Processing.]

3:20

5pSP9. Speech forensics: Automatic acquisition device identification. Daniel Garcia-Romero and Carol Espy-Wilson (Dept. of Elec. Eng., Univ. of Maryland, College Park, MD 20742, dgromero@umd.edu)

Automatic identification-of-the-source is an important problem in the field of speech forensics. That is, given a speech recording, the problem is to determine how to automatically determine the exact make and model of the acquisition device regardless of the speaker and speech content. Recent work developed by the authors have shown that statistical modeling of the device characteristics in terms of MFCCs provides excellent discriminative power when the recordings are obtained under controlled conditions (i.e., all recordings come from the same room and a unique device per class is used). In this study we propose to expand these results by analyzing the effects that room acoustics and intra-device variability have in the source identification accuracy. To perform this analysis, the NIST 2008 SRE dataset will be used. This corpus provides recordings acquired with multiple devices of the same make and model in different rooms.